교육학석사학위논문

온라인 증명 동료평가의 신뢰도 및 타당도 연구

2018년 2월

서울대학교 대학원

수학교육과

오 예 린

온라인 증명 동료평가의 신뢰도 및 타당도 연구

지도교수 권 오 남

이 논문을 교육학 석사학위논문으로 제출함

2017년 11월

서울대학교 대학원

수학교육과

오 예 린

오예린의 석사학위논문을 인준함

2017년 12월

위 원 장 최 영 기 (인)

부위원장 이 경 화 (인)

위 원 권 오 남 (인)

국문초록

온라인 증명 동료평가의 신뢰도 및 타당도 연구

증명학습의 중요성에도 불구하고, 중고등학생 뿐만 아니라 대학생들도 증명 학습에 있어 상당한 어려움을 겪고 있다는 것이 보고되어 왔다. 따라서 특별히 증명을 많이 다루는 대학 전공 수업에서, 학생들의 증명 학습을 위한 새로운 도구의 도입을 검토해볼 만하다. 동료평가(peer-assessment)란 학생들이 평가자가 되어 서로의 과제를 평가하는 활동을 지칭한다. 동료평가는 주로 대학 수준에서 글쓰기나 실습과제의 평가에 활용되어 왔으며, 국내 대학에서도 동료평가 프로그램이 개별 교수의 수업 개선을 가능하게 하고 학과나 대학차원에서의 다양한 효과를 가진다는 것이 확인되었다. 한편, 동료평가가 대학 수준의 수학 수업에서 증명 학습에 사용된 경우는 찾아보기 어렵다.

이처럼 동료평가의 유익이 많음에도 불구하고 왜 증명 동료평가는 실제 대학 수학교육에서 잘 활용되지 않고 있는 것일까? 학생들이 배경지식이나 평가 경험이 부족하기 때문에 학생들의 평가가 신뢰할만한지 확신하지 못하는 것이 가장 큰 이유일 것이다. 그런데 학생들이 수행한 증명 동료평가 결과의 신뢰도나 타당도를 검증한 연구는 아직 없다. 따라서 학생들의 평가에 대한 신뢰도와 타당도를 확인하는 것이 필요하다. 이에 따라 본 연구에서는 증명 동료평가의 신뢰도와 타당도를 살펴본다.

서울 소재 한 대학교의 〈정수론〉 수업에서는 주별 예습과제의 일환으로 동료평가를 실시하였다. 학생들은 매주 수학적 내용을 배우기에 앞서

스스로 증명을 작성하고 제출하도록 요청받았는데, 매주 이 예습과제 문항 중 4개에 대한 동료평가를 실시하였다. 학생들은 정리별로 3명의 동료가 작성한 증명을 채점하였다. 채점을 할 때에 학생들은 교수자가 미리 제공한 채점기준에 따라 증명의 논리성, 명료성, 참신성에 대한 점수를 부여하였다. 작성한 증명을 제출하고, 채점자에게 채점할 증명이 분배되고, 채점자가 채점을 하고, 채점된 결과가 피채점자에게 전달되는 등 증명 동료평가의 모든 과정은 온라인 동료평가 도구인 Classprep을 통해 이루어졌다.

본 연구에서는 먼저 해당 수업의 학생들의 증명 동료평가 결과를 바탕으로 대학생들의 증명 동료평가의 신뢰도를 살펴보았다. 어떠한 평가의 신뢰도는 일치도(equivalence), 일관성(consistency) 그리고 안정성(stability)의 세 가지 측면에서 조사할 수 있다. 본 연구에서는 한 집단에만 동료평가를 실시하였으므로 안정성의 조사는 불가능하였다. 따라서 증명 동료평가의 신뢰도를 조사하기 위하여 안정성을 제외한 일치도와 일관성만을 측정하였다.

다음으로 동료평가의 타당도를 살펴보았다. 본 연구에서는 타당도를 조사하는 방법으로 학생들이 전문가와 유사한 채점을 하는지를 조사하는 방법을 사용하였다. 연구자는 타당도 조사를 위하여 별도로 '증명 평가과제' 두 세트를 설계하여 학생들과 전문가에게 해당 과제들을 수행하도록 요청하였다.

측정 결과, 학생들의 채점 점수 간에 상당히 높은 일치도가 나타났으나 개별 학생이 한 학기에 걸쳐 받은 점수의 일관성은 낮았다. 높은 일치도는 동료평가 시에 어떤 채점자가 배정되는지에 큰 영향 없이 피평가자가 받게 되는 점수가 비슷함을 의미한다. 따라서 동료평가 시에 어떠한 평가자가 배정되는 지에 따라 점수가 달라질 수 있으므로 불공정하다는 지적을 불식시킬 수 있다. 한편 낮은 일관성은 한 학기 동안 받은 개

별 학생들의 동료평가 점수가 일정하지 않고 다소 변화한다는 것을 나타낸다. 한 학기 동안 학생들에게 정수론 내용의 학습과 증명에 대한 학습이 이루어진다는 점을 고려한다면 해당 동료평가에서 일관성이 반드시 확보되어야 하는 것은 아니다. 따라서 낮은 일관성이 해당 동료평가의 신뢰도가 낮다는 것을 의미한다고 보기는 어렵다.

학생들과 전문가의 증명 평가과제 수행을 분석한 결과에 따르면, 학기 중반에는 학생들이 전문가와 비슷한 논리성 채점 수행을 하였지만 명료성과 참신성의 채점 수행은 전문가와 매우 다른 수행을 하였고, 학기 말에 이르러서는 학생들이 논리성, 명료성 그리고 참신성 세 가지 측면 모두 전문가와 유사한 채점 수행을 하게 되었다. 따라서 학기 초반과 중반에는 증명 동료평가의 타당도가 다소 낮았으나 학기말에 이르러서는 증명 동료평가의 타당도가 높았다고 해석할 수 있다. 학생들이 시간이 지나면서 보다 전문가와 유사한 채점을 하게 되었으므로, 학생들의 증명에 관한 관점이 전보다 전문가와 유사해졌음을 추측해볼 수 있다.

주요어: 증명, 동료평가, 신뢰도, 타당도
학　번: 2015-23057

목 차

표 목 차

그 림 목 차

I. 서론

1. 연구의 필요성

증명은 연역적 수학 체계를 유기적으로 이어주는 역할을 하며, 증명의 학습은 대학 수준의 수학 교육에 있어 기본적인 구성 요소이다(Moore, 2016). 증명은 수학과 분리해서 생각할 수 없으며, 수학을 하고 수학적인 의사소통을 하고 기록하는 데에 있어 필수적인 요소이다(Schoenfeld, 1994). 연역적인 활동에 해당하는 증명 활동은 귀납적 방법에 기초한 다른 학문과 수학의 본질적인 차이를 만들어낸다(Dreyfus, 1990).

이러한 증명은 수학 학습의 도구일 뿐만 아니라(Hanna, 1990; Hersh, 1993), 수학교육자들에 의해 수학 교수·학습의 핵심 목표로 여겨져 왔다(NCTM, 2000; RAND Mathematics Study Panel; 2003). 특히, 대학 수준의 수학 학습에서 증명 학습이 핵심 목표가 되어야 하며, 수학적 경험의 일부분으로 증명에 대한 경험이 이루어져야 한다는 점이 강조되어 왔다.

한편, 중고등학생 뿐만 아니라 대학생들도 증명 학습에 있어 상당한 어려움을 겪는다는 것이 보고되어왔다(Harel & Sowder, 2007; Stylianou et al., 2009; Antonini & Mariotti, 2008; Harel & Brown 2008; Moore, 1994; Weber, 2001; Thompson, 1996). Martin과 Harel(1989)이 대학교 4학년 예비교사를 대상으로 한 연구에서 삼분의 일 이상의 학생들이 연역적 논증 뿐 아니라 귀납적 논증 또한 타당하다고 판단하였으며 절반 이상의 학생들이 수학적으로 타당하지 않은 논증을 수학적 증명(proof)이라고 대답하였다. 학생들은 형식적으로도 내용적으로도 증명 검증에 어려움을 겪는 것이다. 또한 Harel과 Sowder(1998)는 대학생들이 증명의 내용보다는 형식에 더 집중한다는 것을 확인하였고, Weber와 Alcock(2004)은 전

문가가 증명을 구성할 때에 의미론적으로 접근하는 반면, 학부 학생들이 증명을 구성하는 과정에서 구문론적으로 접근하는 경향이 있음을 발견하였다.

이처럼 증명학습의 중요성에도 불구하고 많은 학생들이 증명학습에 어려움을 겪고 있다. 이에 따라 증명학습에 어려움을 겪는 학생들을 지원하기 위한 여러 노력이 있어 왔다. 그 중 Smith(2006)의 연구와 Smith 외(2009)의 연구에서는 IBL(inquiry-based learning, 탐구기반학습) 방법을 사용한 수업을 들은 학생들이 전통적인 방법의 수업을 들은 학생들에 비해 증명의 수학적 아이디어에 더 집중한다는 것이 관찰되었다. Smith가 다룬 IBL 방법의 수업에서는 수업 전에 학생들에게 미리 문제와 증명을 해결하도록 하였고 수업 시간에 학생들은 미리 해결한 문제에 대한 발표와 토론에 참여하였다. 즉 Smith의 연구결과를 통해 학생들을 수학적 의사소통 활동에 참여시켰을 때에 의미 있는 증명 학습이 일어났음을 확인할 수 있다. 이는 Weber와 Mejía-Ramos(2014)가 학부 학생들이 수학자들과 비슷한 증명 관념을 갖게 하려면 학생들로 하여금 수학적 의사소통에 참여시켜야 한다고 주장한 것과 연결된다. 수업 시간에 학생들에게 수학적 의사소통에 참여하게 하는 방법으로 발표와 토론이 있다면, 강의실 밖에서 학생들이 할 수 있는 수학적 의사소통은 글로 써진 증명을 서로에게 보내고 받은 증명을 검증하고 평가하는 활동이 있다. 종전에 수학자들이 그러했던 것처럼 각자 자신이 작성한 증명을 동료에게 보내고 그 증명을 받은 사람은 해당 증명이 옳은지, 좋은 증명인지 등을 판단하는 것이다. 다시 말해, 동료평가에 학생들을 참여하게 함으로써 학생들에게 교실 밖에서도 수학적 의사소통에 참여하게 할 수 있다.

동료평가란 학생들이 평가자가 되어 서로의 과제를 평가하는 활동을 지칭한다. 동료평가는 주로 대학 수준에서 글쓰기나 실습과제의 평가에

활용되어 왔으며(배수정, 박주용, 2016), 국내 대학에서도 동료평가 프로그램이 개별 교수의 수업 개선을 가능하게 하고 학과나 대학차원에서의 다양한 효과를 가진다는 것이 확인되었다(신종호, 2014). 반면, 동료평가가 대학 수준의 수학 수업에서 증명 학습에 사용된 경우는 찾아보기 어렵다. 증명을 쓰고 읽는 활동은 수학적 의사소통 과정으로, 증명 쓰기는 일종의 글쓰기 활동이며 증명 읽기는 일종의 읽기 활동이라는 점에서, 증명 학습에서 동료평가의 활용을 고려해볼 만하다.

다양한 글을 읽으며 좋은 글에 대한 인식을 가져가듯이, 다양한 증명을 읽어보는 것은 좋은 증명에 대한 학생들의 관점을 확립시키는 역할을 할 수 있다. 전통적인 대학 수학 수업에서 학생들의 증명 읽기 활동은 교과서 내의 증명을 읽는 것과 또는 교수자가 수업시간에 제시하는 증명을 읽는 것 외에는 별로 없다. 하지만 대학 수학 수업에서 동료평가를 도입하면, 학생들이 수업을 듣는 동료 학생들에게 다양한 사람에 의해 작성된 증명을 읽을 수 있는 기회를 제공할 수 있다.

동료평가 활동의 주요한 특성 중 하나는 학생들로 하여금 평가자가 되어보는 경험을 하게 한다는 것이다. 이러한 경험은 완성된 증명을 읽고 이해하고 기억하는 경험을 넘어서서, 완성된 것인지 아닌지 확실하지 않은 논증을 읽고 그 증명이 완성된 것인지, 증명의 흐름이 잘 읽히는지 등을 평가하는 기회를 제공한다. 즉, 학생들에게 하여금 비판적 증명 읽기의 기회를 제공한다.

이처럼 증명 동료평가의 유익이 많음에도 불구하고 왜 증명 동료평가는 실제 대학 수학교육에서 잘 활용되지 않고 있는 것일까? 학생들이 배경지식이나 평가 경험이 부족하기 때문에 학생들의 평가가 신뢰할만한지 확신하지 못하는 것이 가장 큰 이유일 것이다. 이러한 맥락에서 학생들의 평가 결과를 신뢰하고 사용해도 되는지를 검증하기 위한 노력들이 있

어왔으나(예를 들면, Cho & Schunn, 2007; Stefani, 1994) 학생들이 수행한 증명 동료평가 결과의 신뢰도나 타당도를 검증한 연구는 아직 없다. 따라서 증명 동료평가를 사용하기 전에 학생들의 평가에 대한 신뢰도와 타당도를 확보하는 것이 필요하다. 이에 따라 본 연구에서는 증명 동료평가의 신뢰도와 타당도를 살펴본다. 이처럼 증명 동료평가의 신뢰도와 타당도를 조사함으로써 향후 증명 동료평가의 활용에 대한 함의를 제공할 수 있을 것으로 기대된다.

2. 연구의 목적 및 연구질문

이 연구의 목적은 온라인에서 이루어진 증명 동료평가의 신뢰도와 타당도를 조사함으로써 증명 동료평가의 활용 가능성을 살펴보는 것이다. 구체적으로 학생들이 제출한 평가 내용을 바탕으로 일치도와 일관성의 측면에서 신뢰도를 살펴볼 것이고, 그 다음에는 학생들이 작성한 평가 내용과 전문가 2인이 작성한 평가 내용을 비교함으로써 학생들의 평가가 타당도가 어떠한지 살펴볼 것이다. 이러한 목적을 이루기 위한 구체적인 연구 질문은 다음과 같다.

1. 대학 정수론 수업에서 실시한 온라인 증명 동료평가의 신뢰도는 어떠한가?

2. 대학 정수론 수업에서 실시한 온라인 증명 동료평가의 타당도는 어떠한가?

위 두 가지 연구 질문에 대답하기 위해서 대학 <정수론> 수업에서 온라인 동료평가를 활용한 사례를 분석한다. 연구 질문 1번에 대답하기 위

해서는 학생들이 주별로 수행한 동료평가 결과 및 추가로 수집한 증명 평가과제의 결과를 토대로 학생 채점자간 채점 일치도 및 개별 학생이 받은 점수의 일관성을 조사함으로써 온라인 증명 동료평가의 신뢰도를 조사한다. 연구 질문 2번에 대답하기 위해서는 학생들과 전문가가 수행한 평가 결과가 얼마나 유사한지를 조사함으로써 온라인 증명 동료평가의 타당도를 조사한다.

Ⅱ. 선행 연구 분석

이 장에서는 대학생들의 증명 학습에 관한 선행 연구를 살펴 본 다음 동료평가의 활용에 관한 선행 연구들을 분석한다.

1. 대학생들의 증명 학습

증명 학습에서 동료평가를 도입하는 것의 당위성을 설명하기 위해서는 먼저 대학생들의 증명 학습에 대한 현재까지의 연구를 이해할 필요가 있다. 따라서 이 절에서는 대학생들의 증명 학습에 관련된 선행 연구들을 분석한다.

증명은 연역적 수학 체계를 유기적으로 이어주는 역할을 하며, 증명의 학습은 대학 수준의 수학 교육에 있어 기본적인 구성 요소이다(Moore, 2016). 증명은 수학과 분리해서 생각할 수 없으며, 수학을 하고 수학적인 의사소통을 하고 기록하는 데에 있어 필수적인 요소이다(Schoenfeld, 1994). 연역적인 활동에 해당하는 증명 활동은 귀납적 방법에 기초한 다른 학문과 수학의 본질적인 차이를 만들어낸다(Dreyfus, 1990). 이러한 증명은 수학 학습의 도구일 뿐만 아니라(Hanna, 1990; Hersh, 1993), 수학교육자들에 의해 수학 교수・학습의 핵심 목표로 여겨져 왔다(NCTM, 2000; RAND Mathematics Study Panel; 2003). 특히, 대학 수준의 수학 학습에서 증명 학습이 핵심 목표가 되어야 하며, 수학적 경험의 일부분으로 증명에 대한 경험이 이루어져야 한다는 점이 강조되어 왔다.

한편, 중고등학생 뿐만 아니라 대학생들도 증명 학습에 있어 상당한 어려움을 겪는다는 것이 보고되어왔다(Harel & Sowder, 2007; Stylianou et al., 2009; Antonini & Mariotti, 2008; Harel & Brown, 2008; Moore, 1994; Weber, 2001; Thompson, 1996; Selden & Selden 2008). Harel과 Sowder(1998)는 대학생들이 증명의 내용보다는 형식에 더 집중한다는 것을 확인하였고, Weber와 Alcock(2004)은 전문가는 증명을 구성할 때에 의미론적으로 접근하는 반면, 학부 학생들은 증명을 구성하는 과정에서 구문론적으로 접근하는 경향이 있음을 발견하였다.

현재까지 대학생들의 증명 학습에 관한 다양한 연구가 있었다. 주로 대학생들의 증명 작성에 관한 연구들이 주를 이루고 있는데 Selden과 Selden(2015)은 증명 학습을 다룬 수학교육 문헌들을 범주화하고 서로 다른 범주의 연구 결과들을 연결하는 시도를 하였다. 그들은 증명에 관한 문헌들을 다루는 주제에 따라 증명 독해(proof comprehension), 증명 작성(proof construction), 증명 검증(proof validation) 그리고 증명 평가(proof evaluation)에 관한 문헌들로 범주화하였다. 그런 다음 그들은 해당 범주의 최근 연구들을 소개하고 소개한 연구들 간의 연결을 설명하였다. 이하에서는 대학생들의 증명 학습에 관한 문헌을 Selden과 Selden의 구분에 따라 증명 작성, 증명 독해, 증명 검증 그리고 증명 평가의 네 가지로 범주로 분류하여 분석한 뒤, 네 범주의 연구 결과들이 어떠한 연결을 갖는지를 살펴본다. 마지막으로, 살펴본 연구들이 증명 동료평가에 주는 함의를 살펴본다.

먼저 대학생의 증명 작성에 관련된 문헌들을 살펴본다. 성공적인 증명 작성을 위해 필요한 요소가 무엇인지에 대해 아직 명확히 밝혀진 바는 없으나, 학부 학생들의 증명을 막는 몇몇 방해 요소들이 관찰되었다. 먼저 Selden과 Selden(2008)의 연구에 의하면, 그러한 방해 요소로는 정리 서술의 논리적인 구조를 해석하는 데의 어려움, 존재 양화사와 전칭 양화사 사용에의 어려움, 상징적 기호를 다루는 데의 어려움, 적절한 정보를 알지만 떠올리지 못하는 것, 그리고 어떤 정리들이 중요한지 알지 못

하는 것 등이 있다. 다음으로 비슷한 결과를 나타낸 Weber(2001)의 연구를 살펴보자. 그는 수학 전공 학부생 4명과 박사과정 학생 4명에게 7개의 추상대수 증명 문제를 해결하도록 한 뒤, 그들의 답안을 분석하였다. 그 결과, 간단한 개념만을 이용해서 해결할 수 있는 1, 2번 문항은 모두가 잘 해결하였으나, 특정한 정리를 사용해야 하는 3-7번 문항은 박사과정 학생들은 대부분 올바른 증명에 성공한 반면, 학부 학생들의 정답률은 훨씬 낮았다. Weber는 학부 학생들과 박사과정 학생들 모두 주어진 문제들을 해결하는 데에 필요한 대수학 지식이 충분하다고 가정하기 위해 일부러 추상대수 과목을 수강한 학부 학생들과 대수전공 박사과정 학생들을 연구 대상자로 선정하였다. 그는 해당 문제들을 해결하기 위한 수학적 지식이 두 집단 간에 차이가 별로 없다면, 무엇이 이들의 수행 차이를 야기했는지를 밝히고자 하였다. 이에, 그는 각 집단 참여자들의 증명에서 다음 특징을 보이는 참가자의 수를 조사하였다: 구문론적 지식을 떠올리는 것을 실패함, 구문론적 지식 기초가 불충분함, 논리적 오류. 또한 각 집단 참여자들의 증명에서 관련 없는 추론이 나타난 횟수도 조사하였다. 그 결과, 박사과정 학생들은 세 가지 특징이 거의 나타나지 않은 반면, 학부 학생들은 구문론적 지식을 떠올리는 것에 실패하거나 구문론적 지식 기초가 불충분한 경우가 종종 나타났고 논리적 오류는 나타나지 않았다. 관련 없는 추론이 나타난 횟수는 학부 학생들이 압도적으로 많았다. Weber는 이러한 결과를 통하여 학부 학생들의 전략적 지식의 부족이 그들의 낮은 수행을 야기했다고 주장한다. 즉, 학부 학생들은 필요한 수학적 지식을 알고 있으나, 각 지식을 언제 그리고 어떻게 사용해야할 지에 대한 지식이 부족하다는 것이다.

대학생들의 증명 작성에 관한 다른 연구 결과로는 학부생들에게 증명을 작성하도록 하였을 때에 실험적 논증을 작성하는 경우가 종종 있다는

연구 결과가 있다. Recio와 Godino(2001)는 400명의 대학 신입생에게 대수와 기하 내용의 문장을 증명하도록 하였는데 약 40%의 응답이 실험적 논증이었다. Harel과 Sowder(1998)는 대학 수학 수업을 듣는 학생들을 인터뷰했는데 학생들이 매우 자주 예시를 통해 증명하려 함을 발견하였다. 하지만 Weber와 동료들의 여러 연구에 따르면 도전적인 증명 구성 과제가 주어졌을 때 실험적 논증은 비교적 드물게 나타났다.

이상에서는 대학생들의 증명 작성에 방해를 하는 요소로 전략적 지식의 부족이 있으며 때로 대학생들이 실험적 논증을 작성하기도 한다는 것을 살펴보았다. 위에서 살펴 본 연구들을 비롯하여 학생들에게 증명을 작성해보도록 한 연구는 많지만(예를 들면, Hazzan & Zaskis, 2003; Mamona-Downs & Downs, 2005; Selden & Selden, 2003; Weber, 2008) 학생들이 증명을 어떻게 읽는지를 살펴본 연구는 비교적 드물다(Weber, 2010). 학생들이 실험적 논증을 작성하였다고 해서 그들이 실험적 논증이 설득력 있다고 받아들이는 것이 아닐 수 있음을 밝힌 연구들(Segal, 1999; Healy & Hoyles, 2000)을 통해서 학생들의 증명 작성을 살펴보는 것만으로는 학생들의 증명 학습을 완전히 설명하기 어렵다는 것을 알 수 있다. 이에 따라 이하에서는 대학생들의 증명 독해, 증명 검증 및 증명 평가에 관한 문헌들을 분석한다. 먼저 증명 독해에 관한 연구들을 살펴본다.

대학생들의 증명 독해에 관한 연구들은 크게 독해 전략에 관한 연구들과 독해 능력 측정에 관한 연구들 그리고 인지적으로 접근한 연구들로 분류할 수 있다. 먼저 대학생들의 증명 독해 전략에 관한 대표적인 연구를 살펴보도록 한다. 수학 전공 학부생을 대상으로 한 연구에서 Weber, Brophy와 Lin(2008)은 성공적인 수학전공 학생들은 성공적이지 않은 학생들에 비해서 읽기 전략(정의 재설정하기, 증명하기 전에 명제를 곰곰

이 생각하기 등)을 사용할 가능성이 더 많음을 밝힌 바 있다. 또한 Weber(2015)는 학부 4학년 수학 전공 학생들이 증명의 이해를 촉진하기 위해 사용하는 전략 다섯 가지를 발견하였는데 그 전략들은 다음과 같다.

(1) 증명을 읽기 전에 스스로 증명하려고 시도하기
(2) 주어진 증명에서 사용된 증명 구조 확인하기
(3) 주어진 증명을 부분이나 하위 증명들로 분해하기
(4) 증명의 어려운 주장이 있을 때에는 예시를 통해 실증하기
(5) 주어진 증명에서 쓰인 방법과 자신의 접근을 비교하기

위 연구들은 증명을 읽을 때에 효과적인 전략들이 무엇인지 밝히려고 한 연구들에 해당한다. 전략은 읽기를 할 때에 필요한 도구들이라고 할 수 있는데 읽기를 구성하는 요소들을 밝히고자 한 연구들도 있다. 그 예로 Yang과 Lin(2008)은 증명 독해를 증명을 이해하는 것으로 보았으며, 주어진 증명이 어떻게 작동하는지, 그 증명이 왜 옳은지를 아는 것부터 주어진 증명이 무엇을 증명할 수 있는지를 아는 것까지 포함한다고 하였다. 그들은 증명 독해를 측정하는 도구를 개발하기 위해 증명 독해의 다섯 가지 영역을 작성하였다. 기초지식, 논리적 지위, 요약, 일반화 그리고 적용이 이에 해당한다. ‘기초지식’은 개인이 명제에 포함된 수학적 용어, 그림, 기호와 그 명제의 증명을 이해하는 것을 뜻한다. ‘논리적 지위’는 개인이 증명에서 사용된 주장들의 지위(전제, 결론, 적용된 성질 등)를 인식하는 것을 의미한다. ‘요약’은 개인이 증명의 핵심이나 핵심 아이디어를 식별하는 것을 지칭한다. ‘일반성’은 개인이 명제의 정확도를 올바르게 인식하고 그 증명에 의해 무엇이 정당화 되었는지를 식별하는 것을 말한다. ‘적용’은 개인이 그 명제를 다른 상황에서 적

절하게 적용하는 것을 의미한다. 유효한 증명을 개선시키거나 확장된 질문을 만드는 것 등은 포함되지 않았다.

이처럼 증명 독해의 영역을 범주화 한 연구가 있는 한편 증명 독해력 위한 평가 모델을 제시한 연구도 있다. Mejía-Ramos 외(2012)는 평가 모델을 작성할 때에 증명의 국지적 이해와 전체적 이해를 모두 반영하고자 하였다는 점에서 Yang과 Lin의 연구와 차이가 있다. 그들은 ‘국지적 이해’를 “어떠한 증명을 전체적으로 이해하는 것과는 상반되는 것으로 증명의 용어와 문장들이 무엇을 의미하는지, 어떠한 논리적 지위를 갖는지, 그리고 앞뒤 문장이 어떻게 연결되는지를 아는 것”이라고 규정하였으며 ‘전체적 이해’를 “어떠한 증명을 전체로써 내면화하는 것”이라고 정의하였다. Yang과 Lin이 작성한 독해의 영역 중 ‘기초지식’과 ‘논리적 지위’는 국지적 이해에 해당하고 요약과 일반화는 전체적 이해에 해당한다고 볼 수 있으며 따라서 두 연구에서 제시하는 모델들은 유사한 면이 있다.

이상에서는 대학생들의 증명 독해 전략과 독해 능력의 측정에 대한 문헌들을 살펴보았다. 이상에서 살펴본 증명 독해 연구들은 수학적 논증에 해당하는 증명을 읽는 것이 다른 종류의 글을 읽는 것과 차이가 있다고 간주하고, 증명을 읽을 때에 효과적인 전략들이 무엇인지 증명 독해의 요소들에는 무엇이 있는지를 밝히려고 한 연구들에 해당한다. 이처럼 증명 읽기를 여타 글 읽기와 구분하는 입장을 취하는 연구자들이 있는 반면 증명 읽기를 일종의 글 읽기 활동으로 보는 입장을 취하는 연구자들도 있다. 대학생들이 증명 독해에 관한 문헌 중에는 대학생들이 수학적 증명을 이해하는 데에 어려움을 겪음을 보고한 문헌들이 더러 있는데(예를 들면, Weber & Mejía-Ramos, 2014) Yang(2012)은 대학생들이 증명 이해에 겪는 어려움 중 일부는 독해력과 관련이 있다고 보았다. 그는 고

등학생들의 증명이해의 어려움을 다룬 연구는 대학생들을 대상으로 한 연구에 비해 상대적으로 많지만, 인지적 자료를 통해 설명한 연구는 별로 없음을 지적하였다. 수학교육 분야와 달리, 언어 교육 분야에서는 독해력의 향상에 대한 접근으로 읽기 전략 사용을 강조해왔다. Yang은 증명을 수학적 텍스트의 한 종류로 보고, 언어 교육 분야에서의 접근을 차용하여 증명의 독해력과 읽기 전략 간의 관계를 밝히고자 했다. 그는 증명 읽기 활동에서 적용되는 읽기 전략 중 메타인지적 읽기 전략과 인지적 읽기 전략에 집중하였다[1]. Yang은 533명의 9학년 학생들에게 기하 증명을 읽고 선다형 질문에 대답하게 함으로써 독해력을 조사한 뒤(RCGP 설문), 학생들의 읽기 전략을 조사하는 설문을 실시하였다(CMRS 설문). Yang은 연구결과를 통해 높은 수학적 독해력을 위해서는 적절한 인지적 읽기 전략의 사용과 적절한 메타인지적 읽기 전략 사용이 모두 필요하지만, 적절한 메타인지적 읽기 전략의 사용이 적절한 인지적 읽기 전략의 사용을 이끈다고 보았다.

지금까지 대학생들이 증명을 읽을 때에 어떠한 전략을 사용하는지, 대학생들의 증명 독해 능력을 어떻게 평가할 수 있는지를 다룬 연구들을 살펴보았다. 수학적 증명을 읽는 활동에는 읽기 자료가 타당한지를 따져보는 검증의 과정이 수반된다. 따라서 증명을 읽는 활동을 독해 전략이나 독해력만으로 설명하기에는 다소 무리가 있다. 이에 따라 이하에서는 증명 검증을 다룬 연구들을 살펴본다.

수학교육 문헌들에서 ‘증명 검증’은 시도된 증명의 정부(正否)를 결정하기 위해 증명을 읽고 숙고하는 것을 가리킨다. Selden과

---

1) Pereira-Laird와 Deane(1997)은 메타인지적 읽기 전략은 ‘읽는 활동의 이전, 도중, 이후에 일어나는 전략으로 읽기를 확인하고 해결하기 위한 의도적이고 계획적인 전략’이라고 정의하였고, 인지적 읽기 전략을 ‘읽는 도중에 일어나는 전략으로 독자가 텍스트와 직접적으로 상호작용하는 전략’이라고 정의하였다.

Selden(2003)은 증명을 “정리가 참인 것을 규명하는 글(p. 5)” 이라고 정의하였으며, 증명 검증을 “증명이 참인지를 결정하기 위해 증명을 읽고 반성하는 것(p.5)” 이라고 정의하였다. 본 글에서는 대부분의 수학교육 문헌들에서와 같이 ‘증명 검증’ 을 ‘주어진 증명이 타당한지 판단하는 것’ 으로 규정한다. 대학생들의 증명 검증을 다룬 연구들 중에는 대학생들의 증명 검증 능력이 부족함을 보여주는 연구들도 있지만 이에 반하는 결과를 나타내는 연구들도 있다. 이들을 차례로 살펴보도록 한다.

일찍이 Martin과 Harel(1989)은 대학교 4학년 예비교사를 대상으로 이들에게 친숙한 명제와 친숙하지 않은 명제들에 대한 논증(argument)을 제시하고 해당 논증들이 타당한지 판단하도록 하였다. 주어진 논증 중에는 귀납적 논증도 있었고 연역적 논증도 있었다. 삼분의 일 이상의 학생들이 연역적 논증 뿐 아니라 귀납적 논증 또한 타당하다고 판단하였다. 또한 절반 이상의 학생들이 수학적으로 타당하지 않은 논증을 수학적 증명이라고 대답하였다. 이는 대학에서 수년간 증명 읽기와 증명 작성을 꾸준히 해온 학생들조차도 증명 검증을 제대로 하지 못할 수 있음을 보여준다. 귀납적 논증을 수학적 증명이라고 여긴 것은 수학적 증명이 갖추어야 할 형식에 대한 판단이 잘 이루어지지 않음을, 타당하지 않은 내용의 논증을 타당하다고 여긴 것은 수학적 증명의 내용 타당성에 대한 판단이 잘 이루어지지 않음을 나타낸다. 즉 형식적으로도 내용적으로도 증명 검증에 어려움을 겪는 것이다.

이후로도 비슷한 연구 결과들이 지속적으로 보고되어 왔다. Harel과 Sower(1998)은 학생들이 종종 수학적인 내용보다 형식에 집중하는 ‘외면적 증명 스키마’ 를 가지고 있다고 주장하였다. Inglis와 Mejía-Ramos(2009)는 학생들이 수학적 내용을 통해 논증을 판정할 수 없을 때에 외부적 요인에 의존한다고 보았다. Selden과 Selden(2003)의 연

구에서도 학부 학생들이 증명을 검증할 때에 전체적인 이해에 대한 자신의 느낌을 사용할 때가 많으며 증명의 표면적 특성에 집중하는 경향이 나타났다. 연구자들은 이러한 경향이 학부생들의 증명 검증의 신뢰도를 낮추는 원인이 된다고 보았다. Knuth(2002)는 중등 수학 교사들이 틀린 증명이라도 대수적 조작을 포함하거나 올바른 형식을 가지고 있다면 그 증명이 옳은 증명이라고 받아들인다는 것을 확인하였다. 또한 이 교사들은 실험적 증명이라 할지라도 주어진 명제가 왜 사실인지를 설명한다면 증명으로 받아들였다. 그들이 어떤 수학적 논증의 타당성을 검증할 때에, 해당 논증의 옳고 그름보다는 세부 양식을 충분히 가졌는지 또는 익숙한 증명 기술을 사용하는지를 자주 사용하였다. Knuth는 예비교사들에게 증명을 검토하고 증명에 대해 논의할 수 있는 기회가 거의 주어지지 않음을 지적한다. 그런 뒤 Knuth는 예비교사들에게 증명을 검토하고 증명에 대해 논의할 수 있는 기회를 제공하는 방법으로 Moore 교수법을 지목하였다.

이처럼 대학생들과 수학 교사들이 실험적 논증을 타당한 증명이라고 받아들이기도 한다는 보고가 있어왔으나 최근에 Stylianou, Blanton과 Rotou(2015)가 535명의 학부 학생들을 대상으로 조사한 결과, 대상 학생들의 많은 수가 연역적 증명의 엄밀함과 주된 역할의 진가를 인정하고 인지하고 있었다. 무엇이 이 연구들 간의 서로 다른 연구 결과를 야기했을까? 다시 말해, 어떻게 하면 학생들이 실험적 논증과 연역적 논증의 차이를 인지하고, 연역적 논증의 진가를 인정하게끔 할 수 있을까?

Selden과 Selden(2003)의 연구는 대학생들의 증명 검증 능력 향상의 사례를 보여준다. 그들은 수학 전공 또는 수학교육 전공 학부생 8명을 대상으로 네 개의 수학적 논증을 검증하도록 하였다. 해당 논증들은 정수론 내용의 정리 1개에 관한 서로 다른 논증들로 학부 학생들이 작성한

것이었다. 총 네 번에 걸쳐 주어진 네 개의 논증이 옳은지 여부를 검증하도록 하였다. 처음 두 번의 검증 사이에는 학생들로 하여금 주어진 논증들에 대한 생각을 소리 내어 말하고 증명이 아니라고 생각한다면 그 이유를 말하도록 요청하였다. 마지막 두 번의 검증 사이에는 학생들로 하여금 모든 증명을 함께 읽고 다시 생각해보도록 하였으며 ‘잘 모르겠다.’라고 답변하지 않고 반드시 각 논증이 증명인지 아닌지 여부를 결정하도록 요청하였다. 이들의 연구 결과 학부생들에게 논증을 하나씩 따로 제시하고 검증하도록 한 뒤(46%, 56%) 여러 개의 논증을 한꺼번에 제시하고 다시 검증하도록 하였을 때에(72%, 81%) 올바른 검증을 하는 빈도가 더 높았다. 이러한 결과를 통해서 같은 증명에 대한 논증을 하나씩 제시하는 것보다 여러 개의 논증을 동시에 제시하였을 때에 검증의 정확도가 높아진다고 볼 수 있다. 하지만 단순히 논증을 하나씩 제시하였는지 한꺼번에 제시하였는지의 차이가 검증의 차이를 만들어냈다고 설명하기에는 무리가 있다. 왜냐하면 네 번의 검증 사이에 반성 활동이 들어있었기 때문이다. 비록 검증 사이의 활동들에서 연구자의 직접적인 지시나 안내는 없었으나, Selden과 Selden은 학생들이 보인 검증에서의 향상을 통해 검증 안내가 효과적일 수 있다고 해석하였다. 더불어 이들은 증명 검증을 연습하는 것의 중요성을 강조하며, 특별히 학생들이 작성한 증명에 대한 소그룹 토론이 유익할 것이라고 제언하였다. 특별히 훗날에 자신들의 증명의 옳고 그름을 판단하고 학생들의 새로운 풀이의 옳고 그름을 판단해야 하는 예비 중등교사에게는 특히 그렇다.

위에서 살펴본 ‘증명 검증’은 단순히 증명이 타당한지 또는 옳은지를 판단하는 것을 의미한다. 증명을 적극적으로 읽는 독자는 주어진 증명이 옳은지 검증하는 행위를 할 것이다. 한편 어떤 독자는 이러한 증명 검증 행위뿐만 아니라 주어진 증명의 가치 판단도 할 수 있다. Selden과

Selden(2015)은 증명을 읽는 독자가 하는 가치 판단 행위를 ‘증명 평가’라고 지칭하였으며 이를 ‘증명 검증’과는 배타적인 것으로 정의하였다. 그들은 증명의 옳고 그름을 판단하는 것은 증명 검증으로 지칭하였고, 증명이나 수학적 내용의 가치를 판단하는 것은 증명 평가라고 지칭하였다. 이들은 논지 전개의 편리성을 위해 ‘증명 평가’를 규정할 때에 ‘증명 검증’과 관련된 부분은 배제하고 명료성, 맥락, 설득력, 아름다움, 우아함, 그리고 깊이 등 증명의 특징이나 가치(value)를 판단하는 것만을 증명 평가로 보았다. 이와 다르게 Pfeiffer(2011)는 ‘증명 평가’를 “어떤 증명이 참인지 뿐만 아니라, 그 증명이 명료성, 맥락, 통찰력, 설득력 또는 이해 증진 등의 면에서 얼마나 좋은지를 결정하는 것”이라고 서술하였다. 즉 Pfeiffer는 증명 평가를 증명 검증을 포함하는 개념으로 정의하였다. 연구자는 증명 검증 또한 증명 평가의 일부분으로 받아들인다. 이에 따라 본 글에서는 주어진 증명의 옳고 그름을 판단하는 증명 검증과 주어진 증명의 특징을 살피고 해당 증명에 대한 가치 판단을 내리는 활동을 모두 포괄하여 ‘증명 평가’라고 지칭한다.

증명 작성이나 증명 검증 등에 비해 증명 평가에 관한 연구는 비교적 활발히 이루어지지 않았다. 증명 평가에 관한 연구들을 살펴보면, 증명 평가를 교수자의 교수 실천의 한 부분으로 여기고 접근하였음을 확인할 수 있다. 이러한 이유로 대학생들의 증명 평가에 관한 연구는 거의 찾아보기 어렵다. 따라서 이하에서는 수학자들의 증명 평가를 다룬 연구들을 위주로 증명 평가에 관한 문헌들을 살펴보도록 한다.

먼저 증명 평가에 대한 문헌들을 살펴보기에 앞서 수학적 가치판단에 대해 다룬 문헌을 살펴본다. Tao(2007)는 수학적 내용의 수학적인 질을 판단하는 것은 여러 차원의 판단을 수반한다고 보고 수학자들이 ‘좋은 수학’이라고 말할 때에 ‘좋은 수학’이 무엇을 의미하는지 밝히고자

했다. 그는 좋은 수학적 내용은 이에 뒤따르는 좋은 수학적 발견을 낳는 경향이 있으며, 수학자들은 경험을 통해 어떤 수학적 작업물이 가치가 있는지를 판단할 수 있는 시각을 기르게 된다고 보았다. 그는 '좋은 수학'이 지칭할 수 있는 것 21가지를 제시하였다. 21가지 항목은 '좋은 수학적 문제 해결', '좋은 수학적 기술', '좋은 수학적 이론', '좋은 수학적 통찰', '좋은 수학적 발견' 등으로 구성되어 있다. 즉 Tao는 좋은 수학적 내용을 판단하는 기준을 제시했다기보다 좋은 수학적 내용의 경우를 분류하였다고 볼 수 있다. Inglis와 Aberdein(2014)은 수학적 내용의 가치 판단을 보다 자세히 설명하고자하였다. 이들은 심미성, 복잡성, 유용성 그리고 정확성 네 가지 차원에 걸쳐 수학적 내용의 특성 80가지를 작성하였다.

앞서 살펴본 두 연구에서는 수학적 가치판단의 항목을 각각 21가지 80가지로 제시하는 등 다수의 항목을 제시하고 있는데 Brown과 Michel(2010)은 좋은 수학적 증명의 특징을 가독성, 타당성, 그리고 유창성 세 가지로 제시하였다. 더불어 이들은 이 세 가지 특징(가독성, 타당성, 유창성에 관한 증명 채점 기준을 제시하였다. 후에 Moore(2016)는 이러한 기준에 다른 수학자들도 동의할 수 있는지, 다른 수학자들도 학생들의 증명을 채점할 때에 위 세 가지 기준으로 채점할지에 대해 의문을 가졌다. 그는 네 명의 대학 교수들을 인터뷰하여 학부 학생들이 작성한 증명을 채점해보도록 한 뒤에, 채점 기준을 묻고 좋은 증명의 특징이 무엇인지를 물어보았다. 네 명의 교수 모두가 좋은 증명은 논리적 정확성, 명료성, 유창성, 그리고 증명의 이해에 대한 입증을 지닌다고 대답하였다. Brown과 Michel(2010)의 타당성과 Moore(2016) 연구에서의 논리적 정확성은 논리성에 상응하고, Brown과 Michel(2010)의 가독성과 유창성, Moore(2016) 연구에서의 명료성과 유창성은 명료성에 해당한다고 볼 수

있다.

이상에서 살펴본 연구들은 수학자들의 증명 평가를 다룬 연구였다. 이례적으로 예비수학교사들의 증명 평가를 다룬 연구가 있어 살펴본다. Lavy와 Shriki(2014)는 예비수학교사들이 기하 증명을 위해 만들어 낸 평가 요소들이 어떻게 변화하였는지에 대해 연구하였다. 이 연구에서 예비교사들이 만들어 낸 8가지 평가 요소 중 관례적인 증명 구조, 옳은 해답, 옳은 결론의 형성 그리고 수학적 영역 사이의 연결 지식의 입증은 논리성과 관련이 있으며, 쉽고 이해되는 해답, 문제에 대한 명확한 개요, 그리고 명확한 수학적 기호의 사용은 명료성과 관련이 있고, 합리적인 조직과 명료성은 논리성과 명료성 모두와 관련이 있으며, 마지막에 추가된 요소인 지적 도전성은 참신성과 관련이 있다.

위에서는 증명 작성, 증명 독해, 증명 검증, 증명 평가에 관련된 문헌들을 살펴보았다. 이하에서는 이러한 네 영역이 서로 어떠한 연결을 갖는지 살펴보도록 한다.

‘정의와 정리들을 제대로 알고 제대로 사용하는 것’은 증명 작성과 증명 이해 그리고 증명 검증 모두에서 필요한 요소인데, 차이가 있다면 증명 작성에서는 필요한 정의와 정리들을 마음속에 떠올리는 것이 수반되는 반면 증명 이해나 증명 검증에서는 떠올리는 과정 없이 이미 사용된 정의와 정리가 적절하게 사용되었는지를 판단하기만 하면 된다 (Selden & Selden, 2015). 이러한 측면에서 스스로 증명을 작성하는 것은 다른 이가 작성해 놓은 증명을 이해하거나 검증하는 것에 비해 어려운 일이라고 할 수 있다.

다음으로 증명의 읽기와 작성 간의 관계를 생각해 볼 수 있는데, 증명 읽기와 증명 작성이 서로에게 도움을 준다는 것이 확인되었다. 실제로 독해 관련 연구자들은 읽기와 쓰기를 동시에 활용했을 때에 학생들의 학

습에 긍정적인 효과가 있음을 설명하였다(예를 들면, McGee & Richgels, 1990).

한편 Selden과 Selden(2015)에 의하면 Mejía-Ramos 외(2012)가 정의한 증명 이해와 증명 검증은 여러 가지 공통 요소를 가진다. 문장의 논리적 지위를 점검하는 것, 어떤 증명 구조가 사용되었는지 아는 것, 하위 증명을 작성하는 것, 그리고 증명을 요약하는 것 등이 이에 해당한다. 또한 그들은 이러한 평가를 할 수 있으려면 증명 이해와 증명 작성에 대해 친숙하고 능숙할 필요하다고 주장하였다. 우아함과 같은 특징에 대한 가치 판단을 내리기 위해서는 많은 증명을 보고 작성해 봤어야하기 때문이다. 더불어 그들은 증명 검증을 위해서는 증명 이해가 반드시 선행되어야 한다고 주장하였다. 이를 종합하면 증명 검증과 증명 평가를 위해서는 증명 이해가 반드시 선행되어야 하며, 증명 검증과 증명 평가를 잘하기 위해서는 충분한 증명 읽기와 증명 작성 경험이 필요하다. 이러한 주장은 Selden과 Selden(2003)은 학부생들의 증명 검증을 다룬 연구에서 증명 검증을 위해서는 국지적 이해가 수반되어야 함을 설명한 것으로 뒷받침할 수 있다. 더불어 Pfeiffer(2011)은 학부 학생들을 증명 평가에 참여하게 함으로써 학생들이 증명의 역할을 제대로 인식하게 하고, 그들의 증명 작성에도 도움을 줄 수 있을 것으로 추측하고 이를 검증하고자 하였는데 연구 결과 몇몇 증거들을 찾았다. 이와 같은 맥락에서 Selden과 Selden(2015)은 학부 학생들의 증명 작성 능력을 기르는 것이 그들의 증명 검증 능력의 향상과 직결되는 것이 아님을 발견하고 증명 검증을 따로 명확하게 가르쳐야 한다고 주장하였다.

훗날에 학생들에게 수학적으로 옳은 수학적 내용을 안내하고, 학생들의 수학적 결과물이 옳은지를 판단해야한다는 점에서 예비교사들은 증명 검증 능력을 길러야 한다. 그러나 증명 검증은 대학에서 수학을 공부한

다고 해서 저절로 길러지는 것은 아닌 것으로 보인다. 따라서 예비교사들에게 증명 검증의 경험을 제공하고 잘 안내하는 것이 필요하다. 그러나 전통적인 대학 수학 수업을 통해 대학생들이 접하는 증명은 이미 검증을 마친 것들이 대부분이다. 대학 수업을 통해 학생들이 접하는 증명은 크게 교과서에 실린 증명과 교수자가 제시해주는 증명으로 구분할 수 있다. 교과서의 증명은 이미 검증을 거친 뒤에 인쇄되어 학생들에게 제시된 것이며 교수자가 제시하는 증명은 교수자 스스로 검증을 한 뒤에 학생들에게 제시되기 때문이다. 따라서 전통적인 대학 수학 수업에서 학생들은 증명 검증을 경험할 기회를 거의 갖지 못한다.

한편 대학 수학 전공 수업에서 증명 동료평가의 도입을 통해 학생들에게 증명 검증의 기회를 제공할 수 있다. 또한 스스로 증명을 작성하는 것으로 동료평가가 시작되므로 학생들은 증명 작성의 기회도 갖는다. 증명 검증을 위해서는 증명 독해가 반드시 선행되어야하기 때문에 학생들에게 증명 독해의 기회 또한 주어진다. 또한 제시하는 평가 기준에 따라 증명 검증과 더불어 증명 평가도 이루어질 수 있다. 예를 들어 참신성이나 심미성을 평가 기준에 포함하는 경우, 학생들은 동료가 작성한 증명의 가치 판단을 나름대로 해야 하기 때문이다. 또한 동료평가에서 한 명의 평가자가 평가하는 논증(과제)의 개수를 적절할 수 있는데, 이를 통해 학생들에게 하나의 정리에 대한 여러 개의 논증을 함께 제시할 수 있다. 여러 개의 논증을 함께 제시하고 검증하게 하였을 때에 단일 논증을 검증하게 하였을 때보다 옳은 검증을 한 학생들이 더 많았음을 밝힌 Selden과 Selden(2003)의 결과를 통해, 평가자 1인 당 여러 개의 논증을 평가하도록 하는 것이 증명 동료평가의 신뢰도를 더 높일 수 있을 것이며 학생들의 증명 검증 능력 향상에 더 효과적일 것으로 기대된다.

2. 동료평가

동료평가란 학생들이 평가자가 되어 서로의 과제를 평가하는 활동을 지칭한다. 이 절에서는 동료평가 관련 연구들을 분석한다. 분석을 통하여 동료평가가 어떻게 활용되어 왔는지, 동료평가의 종류에는 어떤 것들이 있는지, 동료평가 활용의 효과나 어려움은 무엇인지 등을 살펴본다. 그 다음, 본 연구에서 사용하는 동료평가는 동료평가의 역사 속에서 어떠한 위치를 가지는지, 증명 동료평가의 기대효과는 무엇인지 등을 설명하도록 한다.

동료평가는 학습자가 동등한 지위(equal-status)를 가진 이들의 생산물이나 수행을 살펴보고 그것의 등급이나 가치 또는 질을 평가하는 것으로(Topping, 2009) 보통 과제물 당 다수의 평가자가 평가하여 교차 채점을 통해 객관성과 신뢰도를 확보하기 위해 사용된다(한국교육평가학회, 2004). 동료평가는 평가의 새로운 형식으로 보일 수 있으나 수백 년 간 사용되어 왔다. 1774년부터 1826년까지 글라스고 대학의 교수로 재직했던 George Jardine은 일찍이 글쓰기에서의 동료평가 방식과 그 이점을 설명한 바 있으며(Gaillet, 1992) 그 후로도 글쓰기 학습에 상당한 효과가 있음이 밝혀져 왔다(O’Donell & Topping, 1998). 동료평가는 그동안 초·중·고등학교 모두에서 성공적으로 활용되어 왔다(Scruggs & Mastropieri, 1998).

동료평가 역시 다른 평가 방식과 마찬가지로 수행 목적에 따라 총괄평가적 목적의 동료평가와 형성 평가적 동료평가로 나눌 수 있는데, 형성적 목적의 동료평가는 과제에 대한 평정에 목적을 두기 보다는 동료평가를 학습자 중심의 학습 방법으로 여기고 평가의 경험을 통하여 학습자의 학습을 개선시키는 것을 목적으로 한다(김민정, 2008). 이러한 형성적

목적의 동료평가는 동료평가 과정을 통해 평가자인 학습자나 피평가자인 학습자 모두에게 인지적, 초인지적 이점을 극대화하여 궁극적으로 학습을 성공적으로 이끄는 것을 목적으로 하고 있다(Kim, 2005).

동료평가가 학생들에게 미치는 긍정적인 효과는 학습적인 측면과 정의적인 측면으로 나누어 볼 수 있다. 먼저 학습적인 측면에서, 학습자는 동료의 과제를 분석하고 오류를 파악하는 등의 인지적, 메타인지적 활동들로 해당 교과 및 주제 영역에 대한 이해를 증진시킬 수 있다(Topping & Ehly, 1998). 또한 정의적인 측면에서, 학습자는 능동적인 학습참여를 통해 자기주도적 학습력을 신장시키게 되며, 평가자로서의 주도권과 주인의식의 경험을 통하여 자아 존중감을 향상시키게 되고, 부정적인 피드백에 대한 두려움을 덜 느끼며, 교수자로부터 피드백을 받을 때에 비해 피드백을 긍정적으로 받아들인다(김정환, 조유미, 2006; Topping et al, 2000; Dochy, Segers, & Sluijsmans, 1999; Hunter & Russ, 1996).

동료평가가 잘 이루어졌을 때 학생들에게만 긍정적인 효과를 미치는 것은 아니다. 동료평가가 잘 이루어졌을 때의 유익 중 하나는 교수자의 채점 부담을 덜어준다는 것이다. 만약 학생들끼리 채점하고 피드백을 제공하는 것이 충분히 신뢰성 있고 타당하게 이루어진다면, 교수자는 채점에 할애하던 시간을 수업 준비 등의 다른 일에 사용할 수 있을 것이다. 교수자의 부담을 덜 뿐만 아니라 궁극적으로는 양질의 교육을 제공하게 될 것이다. 뿐만 아니라 동료평가를 사용하면 학생들은 보다 즉각적인 피드백을 받을 수 있다. 교수자가 학생들의 모든 과제물을 채점하고 피드백을 제공하는 것에 비해, 학생들이 몇몇 동료의 과제물을 채점하는 데에 더 적은 시간이 소요될 것이기 때문이다. 더불어 국내에서 동료평가는 주로 대학 수준에서 글쓰기나 실습과제의 평가에 활용되어 왔으며 (배수정, 박주용, 2016), 국내 대학에서도 동료평가 프로그램이 개별 교수

의 수업 개선을 가능하게 하고 학과나 대학차원에서의 다양한 효과를 가진다는 것이 확인되었다(신종호, 2014).

동료평가는 이와 같은 효과와 장점을 지니지만, 여러 문제점과 우려사항이 지적되어 왔다. 무엇보다 학습자들의 채점에 대한 신뢰성이 확보되지 않았을 때에는 학습자들의 학습에 혼란을 일으킬 수 있다. Orsmond 외(2000)의 연구에서는 부정확한 피드백을 받았을 때에 평가를 받은 학습자는 불편한 감정을 느끼고 동료의 평가 능력에 불신하게 되었음을 확인하였다. 또한 평가자와 피평가자가 누구인지 공개되는 경우에 학생들은 환경적, 심리적으로 안전하지 않다고 느꼈다. 또한 평가자가 자신이 평가하는 과제를 작성한 사람이 누구인지 아는 경우 후광 효과로 인한 평가 결과 오염이 발생할 수 있고 이는 평가의 신뢰도를 떨어뜨리는 원인이 될 수 있다. 뿐만 아니라. 수업 시간에 동료평가를 실시할 때에는 수업 시간 할애의 문제가 있다. 쪽지 시험이나 과제물을 교사가 채점하는 경우에는 수업 이외의 시간을 할애하여 채점할 수 있으나, 학생들이 수업 시간에 동료평가를 실시하는 경우에는 수업 시간을 할애할 수밖에 없다. 특히나 정기적인 동료평가를 실시하고자 하는 경우에는, 교수자가 이러한 수업 시간 할애에 부담을 느낄 수 있다.

웹기반 동료평가의 도입은 이러한 쟁점들을 다소 해결해 준다. 여러 웹기반 동료평가 시스템 중 하나인 Classprep은 채점할 과제물을 무작위로 평가자에게 배정해주고, 채점 결과를 다시 수집하여 피평가자에게 전달해준다. 모든 과정에서 평가자와 피평가자는 서로 알 수 없기 때문에 평가자할 때에 심리적으로 보다 안전하다고 느끼게 될 것이며, 평가하는 과제의 작성자를 알 수 없으므로 후광효과 등을 방지하여 평가 결과 신뢰도를 높일 수 있다. 또한 과제를 제출하고, 동료채점을 하는 모든 과정이 웹상에서 이루어지므로 수업 시간을 할애하지 않고 학생들이 개별

적으로 수업 이외의 시간에 수행하도록 할 수 있다. 더불어 동료평가가 이루어진 이후에 평가자로부터 받은 점수와 피드백에 대한 평가 단계를 추가함으로써, 부정확한 평가가 이루어지는 것을 방지하고자 했다. 또한 평가자와 피평가자가 공개되지 않기 때문에 학생들이 환경적, 심리적으로 안전하지 않다고 느끼는 것을 해결할 수 있다.

이처럼 증명 동료평가의 유익이 많음에도 불구하고 동료평가가 대학 수준의 수학 수업에서 증명 학습에 사용된 경우는 찾아보기 어렵다. 현재 글쓰기 과제에 동료평가가 적극적으로 활용되고 있는데 증명을 쓰고 읽는 활동은 수학적 의사소통 과정으로, 증명 쓰기는 일종의 글쓰기 활동이며 증명 읽기는 일종의 읽기 활동이라는 점에서, 증명 학습에서의 동료평가의 활용을 고려해볼 만하다. 그런데 왜 증명 동료평가는 실제 대학 수학교육에서 잘 활용되지 않고 있는 것일까? 학생들이 배경지식이나 평가 경험이 부족하기 때문에 학생들의 평가가 신뢰할만한지 확신하지 못하는 것이 가장 큰 이유일 것이다. 이러한 맥락에서 학생들의 평가 결과를 신뢰하고 사용해도 되는지를 검증하기 위한 노력들이 있어왔으며 (Cho & Schunn, 2007; Stefani, 1994 등) 특히 글쓰기 과제의 신뢰도와 타당도는 이미 조사된 바 있다(배수정, 박주용, 2016). 반면 학생들이 수행한 증명 동료평가 결과의 신뢰도나 타당도를 검증한 연구는 아직 없다. 동료평가의 신뢰도와 타당도를 조사한 연구의 70% 가량이 동료평가가 충분한 신뢰도와 타당도를 가지고 있음을 확인하였으나(Topping, 2009) 증명 동료평가에서도 신뢰도와 타당도가 높을 것이라고 확언할 수는 없다. 따라서 증명 동료평가를 사용하기에 앞서 학생들의 평가에 대한 신뢰도와 타당도를 확보하는 것이 필요하다. 이에 따라 본 연구에서는 증명 동료평가의 신뢰도와 타당도를 살펴본다.

증명 동료평가의 신뢰도와 타당도를 조사하는 방법을 구체적으로 설정

하기에 앞서 신뢰도와 타당도 관련 문헌들을 살펴보는 것이 필요하다. 따라서 이하에서는 동료 평가의 신뢰도와 타당도 관련 문헌들을 살펴본다. Cohen, Manion과 Morrison(2011)은 평가의 신뢰도를 검증하는 방법으로는 세 가지 방법으로 일치도(equivalence) 검증, 일관성(consistency) 검증, 그리고 안정성(stability) 검증을 제시하였다. 여기서 일치도는 같은 시점에 한 개인의 특정 능력을 측정(평가)하였을 때에 측정된 결과들이 얼마나 유사한가를 의미한다. 일치도를 확보하는 방법은 두 가지가 있는데 첫 번째 방법은 평가 시에 동일한 형식을 사용하는 것이고, 두 번째 방법은 채점자간 일치도를 확보하는 것이다. 본 연구에서는 주별로 각 학생들이 수행한 증명 과제가 동일하게 함으로써 일치도 확보를 하였으므로, 채점자간 일치도 확보만이 남아 있다. 따라서 본 연구에서는 두 번째 측면의 일치도 검증을 위해, 주별로 실시한 동료평가에서 동일한 증명에 대한 채점한 세 명의 평가가 얼마나 유사한지를 검사하도록 한다. 일관성은 다른 시점에 수집된 점수들 간의 유사성을 의미한다. 보통 일관성을 조사할 때에는 같은 내용의 평가를 두 번 실시한 후에 평가 결과가 얼마나 유사한지를 살펴보는 방법을 따른다. 본 연구에서는 학생들이 11주에 걸쳐 실시한 동료평가 결과를 바탕으로 개별 학생이 받은 점수의 내적 일관성을 조사하는 방법으로 일관성 검증을 실시한다. 안정성은 서로 다른 시간에, 서로 다른 집단에 동일한 측정 및 평가를 실시하였을 때에 유사한 결과를 얻을 수 있는지와 관련된 것이다. 본 연구에서는 하나의 집단만을 대상으로 하므로 안정성 검증은 실시할 수 없었다.

또한 어떤 평가의 타당도란 “검사도구가 측정하고자 하는 능력이나 특성을 얼마나 충실하게 재고 있는가”를 의미한다(Gronlund & Linn, 1990). 평가의 타당도를 확보하는 방법으로 내용 타당도, 기준 타당도, 구인 타당도, 안면 타당도, 예언 타당도 등을 확보・검증하는 방법이 있

다(Cohen, Manion & Morrison, 2011). 본 연구에서는 증명 동료평가의 타당도를 검증하는 방법으로 전문가의 평가와 학생들의 평가 간의 일치도[2]를 조사하였다. 학생들이 전문가와 비슷한 평가를 하는지를 조사함으로써 증명 동료평가를 통해 타당한 평가가 이루어지고 있는지를 검사할 수 있기 때문이다.

이처럼 증명 동료평가의 신뢰도와 타당도를 조사함으로써 향후 증명 동료평가의 활용에 대한 함의를 제공할 수 있을 것으로 기대된다. 본 연구에서는 증명 동료평가의 신뢰도와 타당도를 검증할 때에 크론바흐 알파 방법(Cronbach, 1951)을 사용하였다. 이 방법은 가장 널리 사용되고 있는 신뢰도 측정 방법으로(Tavakol & Dennick, 2011) 내적일관성으로써 신뢰도를 계산하는 방법 중 하나이다. 이때에 계산되는 계수를 크론바흐 알파 계수(Cronbach’s alpha coefficient) 또는 신뢰도의 알파 계수(alpha coefficient of reliability)라고 부른다. 이 계수는 아이템들 간의 내적 일관성의 척도이며 다중항목 스케일(multi-item scales)에서 사용된다(Santos, 1999). 크론바흐 알파 계수는 보통 0에서 1사이의 값을 가지는데 때때로 음의 값을 가지기도 한다. 용인할 수 있는(acceptable) 신뢰도 수준에 대한 기준은 연구자 마다 약간씩 다른데 보통 신뢰도가 0.67 이상이면 용인할 수 있는 수준으로 취급한다. Nunnaly(1978)은 알파 계수가 0.7이 넘으면 용인할 수 있는 신뢰도라고 보았고 Bryman과 Cramer(1990)은 0.8 이상이면 용인할 수 있는 신뢰도라고 주장하였다. 최근에 와서는 알파 계수를 통한 신뢰도 판단 기준의 세분화가 이루어졌다. 대표적으로 Cohen, Manion과 Morrison(2011)은 알파 계수 측정을 통한 신뢰도 검증의 가이드라인을 <표 Ⅱ-1>과 같이 제시하였다.

---

2) Topping(2009)은 이러한 일치도를 ‘동료평가의 정확성(accuracy of peer assessment)’이라고 명명하였다.

<표 Ⅱ-1> 알파 계수 측정을 통한 신뢰도 검증의 가이드라인 (Cohen, Manion & Morrison, 2011, p. 640)

크론바흐 알파 계수	신뢰도
0.90 이상	신뢰도 아주 높음 (Very highly reliable)
0.80 이상 0.90 미만	신뢰도 높음 (highly reliable)
0.70 이상 0.80 미만	신뢰할 만 함 (reliable)
0.60 이상 0.70 미만	미미하게 신뢰할 만 함 (marginally/minimally reliable)
0.60 미만	용납할 수 없게 신뢰도 낮음 (unacceptably low reliability)

Cohen, Manion과 Morrison은 신뢰도의 정도를 신뢰도 아주 높음, 신뢰도 높음, 신뢰할 만 함, 미미하게 신뢰할 만 함 그리고 용납할 수 없게 신뢰도 낮음의 다섯 가지 단계로 분류하였다. Nannaly나 Bryman과 Cramer가 용납할 수 있는 신뢰도인지 아닌지를 결정하는 기준을 제시한 데에 그친 것과는 차이가 있다. 본 연구에서 알파 계수 측정을 통해 신뢰도와 타당도를 계산할 때에는 평가의 신뢰도와 타당도 검증에서 널리 사용되고 있는 Cohen, Manion과 Morrison의 가이드라인을 따른다.

Ⅲ. 연구방법

1. 연구 참여자

1.1. 대학생 참여자

본 연구는 수도권 소재 대학교 <정수론> 과목의 수강생 25명 중 연구 참여를 희망한 학생들을 대상으로 이루어졌다. 본 과목의 수강생들은 해당 과목을 수강 신청하기 이전에 공지된 강의계획을 통해 해당 수업에서 주별과제로 동료평가가 이루어질 것임을 알고 있었다. 수강생 25명 중 18명은 수학교육 전공 학생이었고 5명은 과학교육 전공 학생이었으며 나머지 2명은 인문계열 전공 학생이었다. 본 과목은 학부 2학년 대상 과목이었고, 수학교육 전공 수강생 중 대부분이 2학년 학생들이었다. 또한 수강생 25명은 여학생 2명, 남학생 23명으로 이루어져 있었다. 본 연구에서는 자료를 수집할 때마다 따로 수강생들에게 연구 참여 의사를 물어보았는데, 이 때문에 수집된 자료별 참여자 수가 상이하다. 수집된 자료별 참여자 수는 다음 <표 Ⅲ-1>와 같다.

<표 Ⅲ-1> 자료별 대학생 연구 참여자 수

자료 분류	참여자 수
온라인 증명 동료평가	16명
증명 평가과제 1	20명
증명 평가과제 2	14명

학생들에게 학기말에 온라인 설문 형식으로, 본인의 동료평가 활동 자료를 연구에 활용하는 것에 동의하는지를 조사하였다. 수강생 25명 중 총 16명의 학생이 동의한다고 응답하였다. 추후에 신뢰도를 구하는 과정에서는 본 설문에서 동의한다고 응답한 학생들의 동료평가 활동 자료만 활용되었다. 타당도 조사를 위해 실시한 '증명 평가과제 1'과 '증명 평가과제 2'에는 각각 20명, 14명이 참여하였다. 연구 참여에 대한 강제성을 배제하기 위해 각 과제는 익명으로 진행되었다.

1.2. 전문가 참여자

증명 동료평가의 타당도를 조사하기 위하여 전문가 2인에게 정수론 내용에 해당하는 증명 10개의 채점을 의뢰하였다. 이들이 채점한 증명은 학생들이 참여한 '증명 평가과제 1'과 '증명 평가과제 2'의 내용과 동일하다. 본 연구에서는 정수론을 큰 틀에서 대수학으로 보고 대수학을 전공하는 박사과정 학생들을 섭외하였다. 연구에 참여한 전문가 2인 중 1인은 대수학 박사 학위 소지자로 본 연구에 참여하였을 때에는 박사 학위를 취득한 직후였다. 나머지 1인은 대수학 전공 박사과정 학생으로 본 연구에 참여하였을 때에는 박사과정을 시작한지 3년 정도 지난 시점이었다.

2. 동료평가 설계 및 절차

서울 소재의 한 대학교의 <정수론> 수업에서는 수강생들에게 예습을 위한 과제로써 서면과제와 온라인 동료평가를 수행하도록 하였다. 해당 수업은 학생들의 발표와 토론으로 진행되었는데 학생들은 수업에 앞서 미리 배정된 정리들의 증명을 스스로 작성하도록 요청받았다. 학생들에

게 배정된 정리 중 주별로 4개의 정리가 동료평가에 활용되었다. 동료평가 시에 학생들은 정리별로 3명의 동료가 작성한 증명을 채점하였다. 채점을 할 때에 학생들은 교수자가 미리 제공한 채점기준에 따라 증명의 논리성, 명료성, 참신성에 대한 점수를 부여하였다. 작성한 증명을 제출하고, 채점자에게 채점할 증명이 분배되고, 평가자가 평가를 하고, 평가된 결과가 피평가자에게 전달되는 등 증명 동료평가의 모든 과정은 온라인 동료평가 도구인 Classprep을 통해 이루어졌다. 이 절에서는 <정수론> 수업에서 실시한 증명 동료평가를 어떻게 설계하였는지를 설명한 뒤에 자료 수집 절차에 대해 설명한다.

2.1. 설계

2.1.1. 증명 채점기준

연구자는 학생들이 증명 동료평가 시에 활용할 수 있는 채점기준을 작성하였다. 매주 서로 다른 4개의 정리에 대한 증명들을 채점해야 한다는 점을 감안하여 증명의 채점에 활용될 수 있는 일반적인 채점기준을 작성하여 학생들에게 제시하는 것이 효율적일 것으로 판단하고 어떤 증명의 채점에든 일반적으로 사용할 수 있는 채점기준을 설계하였다.

배수정과 박주용(2016)의 연구에서는 심리학과 학부 전공 수업에서 실시한 글쓰기 과제 동료평가 사례를 두 가지 설명하였는데, 첫 번째 사례에서는 글쓰기 동료평가의 채점기준으로 통찰, 논리, 흐름의 세 가지 차원이 사용되었고, 두 번째 사례에서는 통찰과 글쓰기 두 가지 차원이 사용되었다. 그러나 증명 쓰기는 특정한 장르의 글쓰기이기 때문에(Selden & Selden, 2013), 증명 동료평가를 위한 별도의 채점기준이 필요하다.

증명 학습의 중요성과 학생들이 증명 학습에서 겪는 어려움에 대한 연

구는 많이 이루어져 왔으나, 이에 비해 증명 채점에 관한 연구는 활발히 이루어지지 않았다. 현재까지 증명 채점에 관한 연구들은, 증명 채점을 교수자의 교수 실천(teaching practice)의 한 부분으로 여기고 접근하였다는 점에서 본 연구의 목적과는 차이가 있으나 이를 바탕으로 채점 항목 및 기준을 작성하는 데에는 무리가 없었다. 선행연구들에서 밝힌 채점 항목과 본 연구에서 사용한 증명 채점 항목을 표로 정리하면 다음 <표 Ⅲ-2>와 같다.

<표 Ⅲ-2> 선행연구 및 본 연구의 증명 채점 항목

Brown과 Michel (2010)	Moore (2016)	Lavy와 Shriki (2014)	본 연구
타당성	논리적 정확성	관례적인 증명 구조, 옳은 해답, 옳은 결론의 형성, 수학적 영역 사이의 연결 지식의 입증	논리성
가독성, 유창성	명료성, 유창성	쉽고 이해되는 해답, 문제에 대한 명확한 개요, 명확한 수학적 기호의 사용, 합리적인 조직과 명료성	명료성
		지적 도전성	참신성

먼저 Brown과 Michel(2010)은 좋은 수학적 증명의 특징을 가독성, 타당성, 그리고 유창성 세 가지로 제시하였으며, 이 세 가지 특징(가독성, 타당성, 유창성)에 관한 증명 채점 기준을 제시하였다. Moore(2016)는 이러한 기준에 다른 수학자들도 동의할 수 있는지, 다른 수학자들도 학생들

의 증명을 채점할 때에 위 세 가지 기준으로 채점할지에 대해 의문을 가졌다. 그는 네 명의 대학 교수들을 인터뷰하여 학부 학생들이 작성한 증명을 채점해보도록 한 뒤에, 채점 기준을 묻고 좋은 증명의 특징이 무엇인지를 물어보았다. 네 명의 교수 모두가 좋은 증명은 논리적 정확성, 명료성, 유창성, 그리고 증명의 이해에 대한 입증을 지닌다고 대답하였다.

Brown과 Michel(2010)의 타당성과 Moore(2016) 연구에서의 논리적 정확성은 논리성에 상응하고, Brown과 Michel(2010)의 가독성과 유창성, Moore 연구에서의 명료성과 유창성은 명료성에 해당한다고 볼 수 있다.

Lavy와 Shriki(2014)는 예비수학교사들이 기하 증명을 위해 만들어 낸 평가 요소들이 어떻게 변화하였는지에 대해 연구하였다. 이 연구에서 예비교사들이 만들어 낸 8가지 평가 요소 중 관례적인 증명 구조, 옳은 해답, 옳은 결론의 형성 그리고 수학적 영역 사이의 연결 지식의 입증은 논리성과 관련이 있으며, 쉽고 이해되는 해답, 문제에 대한 명확한 개요, 그리고 명확한 수학적 기호의 사용은 명료성과 관련이 있고, 합리적인 조직과 명료성은 논리성과 명료성 모두와 관련이 있으며, 마지막에 추가된 요소인 지적 도전성은 참신성과 관련이 있다.

위의 연구들을 정리해보면, Brown과 Michel(2010)은 좋은 수학적 증명의 특징으로 논리성과 명료성을 들었고, Moore(2016)의 연구에 참여한 수학과 교수들은 논리성과 명료성을 증명 채점의 주요 기준으로 삼았다. Lavy와 Shriki(2014)의 연구에 참여한 예비교사들 또한 논리성과 명료성을 증명평가의 주요 요소로 삼았으나 연구의 마지막에 이르러서 참신성도 하나의 평가 요소로 삼았다는 데에 차이를 보였다. 본 연구에서는 세 개의 선행 연구를 바탕으로 증명 동료평가에 사용할 채점 항목을 논리성, 명료성, 참신성 세 가지로 정하였다.

연구자와 교수자는 논리성, 명료성과 참신성을 채점 항목으로 정한 뒤에 채점기준을 작성하였다. 작성된 채점기준은 다음 <표 Ⅲ-3>과 같다.

<표 Ⅲ-3> 증명 동료평가를 위한 채점기준

영역	점수	설명
논리성	5점	논리적인 비약이나 오류가 없으며 증명이 완성되었다.
	3점	증명의 군데군데에 논리적인 비약 또는 논리적인 오류가 있다. 증명이 절반 정도 밖에 완성되지 않았다.
	1점	논리적인 비약이나 오류가 심하다. 또는 증명이 거의 완성되지 않았다.
명료성	3점	증명의 구성이 깔끔하다. 수학적 표현이 매끄럽다.
	2점	증명에 필요 없는 내용이 조금 들어있다. 수학적 표현이 약간 매끄럽지 않다.
	1점	증명에 필요 없는 내용이 많다. 수학적 표현이 매끄럽지 않다.
참신성	2점	참신하다.
	1점	그저 그렇다.

연구자와 교수자는 앞서 살펴본 선행연구들에서 논리성과 명료성이 참신성에 비해 많이 강조되고 있기 때문에 논리성과 명료성의 배점을 참신성의 배점보다 높게 잡았으며, 논리적으로 결함이 없어야 옳은 증명이 된다는 점에서 논리성 점수를 명료성 점수에 비해 더 많이 배정하기로 결정하였다. 이에 따라 논리성, 명료성, 참신성의 배점을 각각 5점, 3점, 2점으로 설정하였다. Selden과 Selden(2015)의 분류에 의하면 논리성의 평가는 증명 검증에 해당하고, 명료성과 참신성의 평가는 증명 평가에

해당한다.

2.1.2. 온라인 동료평가 시스템 Classprep

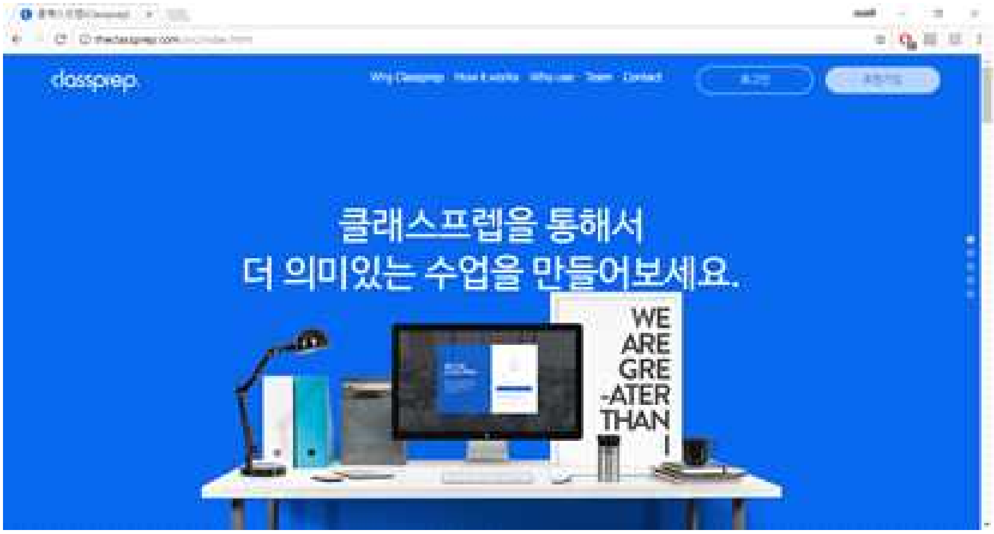
작성한 증명을 제출하고, 채점자에게 채점할 증명이 분배되고, 채점자가 채점을 하고, 채점된 결과가 피평가자에게 전달되는 등 증명 동료평가의 모든 과정은 온라인 동료평가 도구인 Classprep[3)]을 통해 이루어졌다. Classprep은 Park(2017) 이 개발한 웹기반 동료평가 시스템으로 예습을 위한 글쓰기 과제의 동료평가에 활용할 목적으로 개발되었다. 현재까지 국내대학 200여개의 강좌에서 사용되어 왔다. Classprep에서는 누구든지 가입 후에 강좌를 개설할 수 있으며, 학생들이 개설된 강좌의 코드를 등록하면 이후의 모든 동료평가 과정에 참여할 수 있다. 해당 강좌의 페이지에서 과제를 등록 및 수정할 수 있다. 과제 등록 시에는 과제 하나당 평가자 수, 과제 일정, 평가 항목 및 채점점수 스케일 등을 설정할 수 있다. 또한 학생들에게 과제를 하면서 생겼던 질문을 입력하도록 하는 기능도 있다. 강의자가 과제를 등록한 뒤에, 학생들은 다음 세 가지 절차에 따라 동료평가를 수행하게 된다.

(1) 1단계 - 과제 제출
(2) 2단계 - 동료평가 및 자기평가
(3) 3단계 - 피드백 평가

학생들은 먼저 자신이 작성한 글(증명)을 업로드한다(1단계 - 과제 제출). 그 다음 자신에게 배정된 학생들의 글(증명)을 평가한 뒤 자신이 작성한 글(증명)을 평가한다(2단계 - 동료평가 및 자기평가). 마지막으로

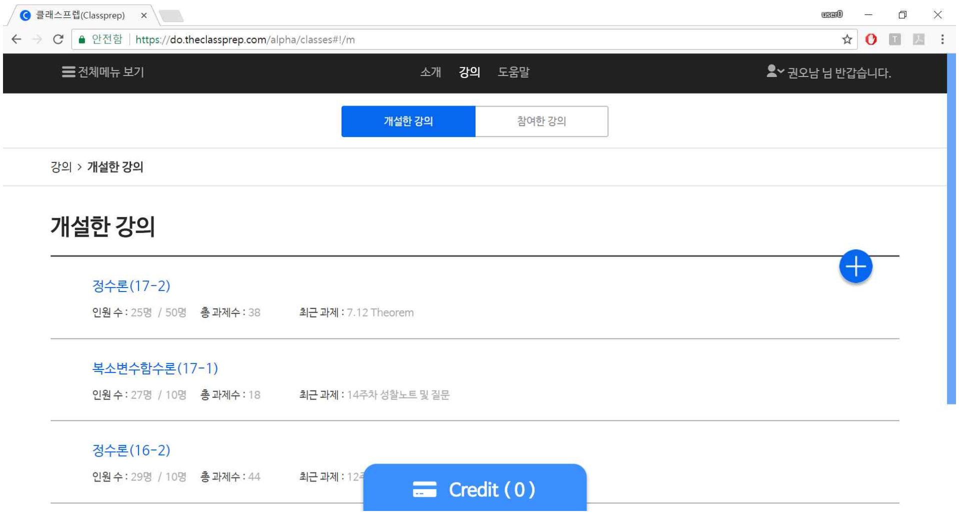
3) http://theclassprep.com/src/index.html

동료들이 자신의 글(증명)에 대해 작성한 평가 내용을 보고 납득할 만한 평가인지, 피드백이 도움이 되었는지 등을 기준으로 동료의 평가에 대한 평가를 한다(3단계 - 피드백 평가). 이 과정에서 학생들은 자신이 평가하는 글(증명)이 누구의 글(증명)인지 알 수 없었으며, 자신의 글(증명)을 평가한 사람이 누구인지도 알 수 없었다. 다음 [그림 Ⅲ-1]은 Classprep 사이트의 메인 페이지이다.



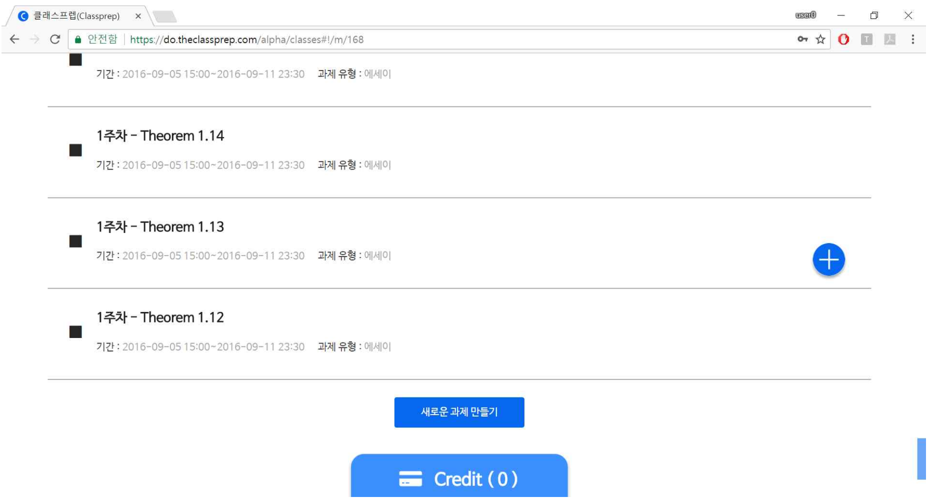
[그림 Ⅲ-1] Classprep 사이트 - 메인 페이지

누구든지 Classprep 사이트에 접속하여 회원가입이 가능하며, 회원가입 후에는 강의 개설 및 강의 참여 신청이 가능하다. 다음 [그림 Ⅲ-2]는 강의 개설 후에 개설된 강의 목록 페이지이다.



[그림 Ⅲ-2] Classprep 사이트 – 강의 목록 페이지 (교수자 화면)

누구든지 로그인 후에 강의를 개설할 수 있으며 강의 개설과 동시에 강의 코드가 부여된다. 강의에 참여할 때에는 로그인 후에 강의코드를 입력하여 개설된 강의에 참여 신청을 할 수 있다. 다음 [그림 Ⅲ-3]은 개설된 강의의 과제 목록 페이지이다.



[그림 Ⅲ-3] Classprep 사이트 – 과제 목록 페이지 (교수자 화면)

강의 개설자와 강의 참여자 모두 과제 목록에서 현재까지 등록된 모든 과제를 확인할 수 있다. 미리보기 화면에서 과제의 기간과 과제 유형을 확인할 수 있다.

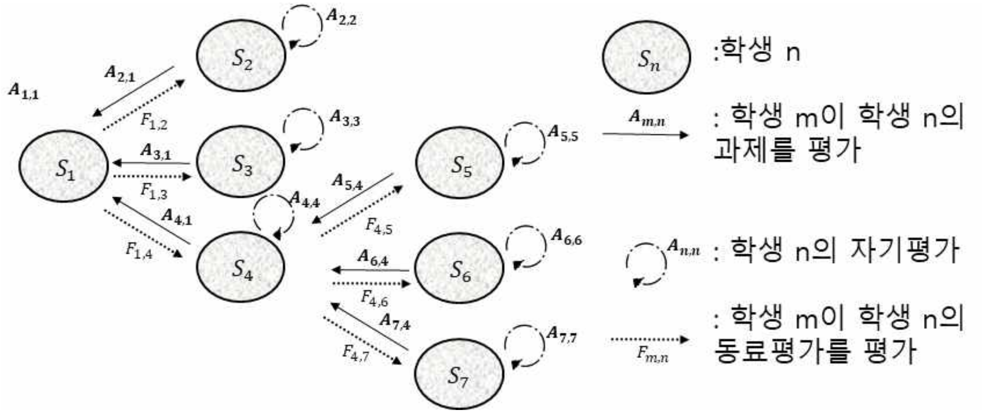
2.2. 절차

본 연구에서 대상으로 한 <정수론> 과목은 변형 Moore 교수법(MMM, modified Moore method)을 기반으로 설계되고 진행되었다. Moore 교수법은 R. L. Moore(1882-1974)에 의해 시작된 교수법으로 학생들로 하여금 의미 있는 수학적 활동에 참여하게 하는 방법으로 고안된 교수법 중 하나다. Moore 교수법의 가장 큰 특징은 교수자가 강의를 하지 않는다는 것이다. Moore는 교수자가 강의를 통해 완성된 증명을 제시하는 기존의 강의 방식을 벗어나 학생들에게 증명해야할 정리의 목록을 학생들에게 제시하고 학생들이 스스로 옳은 증명을 구성하도록 요청하였다(Cohen, 1982). Moore의 수업을 들은 학생들은 수업 이외의 시간에 탐구학습을 하고 수업 시간에는 자신이 발견한 증명에 대해 발표하였다. Moore 교수법의 효과는 그가 배출한 제자들의 업적을 통해 입증되고 있다(권오남, 2009). Moore 이후에 현재까지 Moore의 제자들을 주축으로 여러 형태의 변형 Moore 교수법이 활용되고 있다. 이들 변형 Moore 교수법은 약간씩 차이가 있으나 학생들에게 적합한 문제와 학습 자료를 제시함으로써 학생들 스스로 발견학습을 하도록 하며, 학생들에게 발견, 발표, 그리고 토론의 기회를 충분히 제공하고자 한다는 점에서 공통점을 가진다(Coppin et al., 2009). 해당 <정수론> 강좌의 학생들은 수업에 오기 전에 미리 제공된 목록의 정리들에 대한 증명을 작성해 왔으며, 강의 시간은 주로 학생들이 자신이 작성한 증명을 공유하고 공유된 증명들에 대해 토론하는 데에 사용되었다. 필요한 때에는 교수자가 제시하는 주제를 가지고 모둠

토론을 진행했다.

해당 <정수론> 수업의 교수자는 학생들에게 주별로 15개가량의 정리의 증명을 작성하는 예습과제를 내주었다. 이에 따라 학생들은 매주 초에 그 주에 다룰 정리들의 증명을 손으로 작성하여 제출하였다. 서면으로 제출하는 예습과제와 더불어 2.1절에서 설명한 방식으로 온라인 동료평가가 이루어졌다. 주별로 배정된 15개가량의 증명들 중에 4개의 정리가 동료평가 과제로 지정되었으며, 학생들은 수업에서 해당 정리들을 다루기 직전 주에 Classprep을 통해 증명 동료평가를 수행하였다.[4] 학생들은 각 정리별로 배정된 3개의 증명을 평가하였다. 주별 동료평가는 학기 전체 15주 가운데 중간고사와 기말고사 기간을 제외한 11주에 걸쳐 이루어졌다.

과제 한 개에 해당하는 동료평가의 과정을 한 학생을 중심으로 구조적으로 도식화하면 다음의 [그림 Ⅲ-4]와 같다.



[그림 Ⅲ-4] Classprep 활동의 구조적 설명

4) 학생들이 주별로 증명 동료평가를 실시한 정리들의 목록은 부록에 수록하였다.

[그림 Ⅲ-4]는 학생 4를 기준으로 동료평가의 구조를 나타낸 것이다. 학생 4는 학생 1의 글(증명)을 평가하고($A_{4,1}$; 2단계 - 동료평가 및 자기평가) 학생 1은 학생 4의 동료평가에 대한 평가를 한다($F_{1,4}$; 3단계 - 피드백 평가). 또한 학생 4는 동료 자신의 글(증명)을 평가한다($A_{4,4}$; 2단계 - 동료평가 및 자기평가). 한편, 학생 5, 학생 6, 학생 7은 학생 4의 과제를 평가하고($A_{5,4}$, $A_{6,4}$, $A_{7,4}$; 2단계- 동료평가 및 자기평가) 학생 4는 학생 5, 학생 6, 학생 7의 동료평가에 대한 평가를 한다($F_{4,5}$, $F_{4,6}$, $F_{4,7}$; 3단계 - 피드백 평가). [그림 Ⅲ-4]에서는 동료평가 활동의 모든 구조를 담기보다는 한 학생(학생 4)이 채점자인 동시에 피평가자인 것을 설명하고자 하였다.

본격적으로 증명 동료평가를 실시하기에 앞서, 학기 초 수업 시간에 증명 동료평가 활동에 대한 설명과 채점 기준에 대한 논의가 이루어졌다. 교수자는 먼저 증명 동료평가 활동의 과정 및 채점기준에 대해 설명하였다. 이후에 교수자는 학생들에게 예시 증명을 준 뒤 주어진 채점기준으로 채점한다면 어떻게 채점할 것인지에 대해 고민할 시간을 주었다. 그 다음, 학생들은 어떻게 채점할 것인지에 관해서 자신들의 의견을 나누고 논의하는 시간을 가졌다. 이러한 과정에서 학생들 간에 채점기준에 대한 어느 정도의 합의가 이루어진 것으로 보였다.

3. 증명 평가과제 설계 및 절차

학생들이 전문가와 유사한 평가를 하는지 조사함으로써 학부 수업에서 실시하는 증명 동료평가의 타당도를 조사할 수 있다. 이에 따라, 연구자는 타당도 조사를 위하여 별도로 ‘증명 평가과제’ 두 세트를 설계하여 학생들과 전문가에게 해당 과제들을 수행하도록 요청하였다. 이 절에서

는 ‘증명 평가과제 1’과 ‘증명 평가과제 2’를 어떻게 설계하였는지를 설명한 뒤 자료 수집의 절차에 대해 설명하도록 한다.

3.1. 설계

증명 평가과제 1, 2는 서로 다른 집단 간의 평가 유사도를 측정함으로써 증명 동료평가의 타당도를 조사하기 위해 설계하였다. 각각의 증명 평가과제를 하나의 정리에 대한 서로 다른 5가지의 증명으로 구성하기 위해 먼저 각각의 증명 평가과제에서 사용할 정리(theorem)를 선정해야 했다. 정리의 선정 기준은 다음 두 가지와 같았다.

(1) 수업시간에 비중 있게 다루어진 정리

(2) 다양한 증명이 가능한 정리

먼저, 학생들의 지식 부족이 평가 결과에 영향을 미치지 않도록 하기 위해 수업 시간에 비중 있게 다룬 정리들 중 하나를 선정하고자 하였다. 또한 증명 평가과제에서 다양한 증명을 제시할 수 있으려면 다양한 방법의 증명이 가능한 정리를 선정할 필요가 있었다. 이러한 기준을 토대로 ‘증명 평가과제 1’을 위해서는 ‘소수의 무한성’에 관한 정리가, ‘증명 평가과제 2’를 위해서는 ‘페르마의 소정리’가 선정되었다. 증명 평가과제에서 학생들에게 주어진 ‘소수의 무한성 정리(Infinitude of Prime Theorem)’와 ‘페르마의 소정리(Fermat’s Little Theorem)’의 구체적인 서술은 다음 [그림 Ⅲ-5]와 같다.

증명 평가과제 1: (Infinitude of Prime Theorem)

There are infinitely many prime numbers.

증명 평가과제 2: (Fermat's Little Theorem)

If p is a prime and a is an integer relatively prime to p, then $a^{p-1} \equiv 1 (\mathrm{mod}\ p)$.

[그림 Ⅲ-5] 증명 평가과제에서 다룬 정리

소수의 무한성에 관한 정리와 페르마의 소정리는 각각 학기의 전반부와 후반부에 다루어진 정리들 중 핵심 내용에 해당하며 여러 가지 방법으로 증명이 가능하다. 따라서 이 정리들에 대한 증명을 평가하게 함으로써 학생들에게 <정수론> 과목의 핵심 내용에 관해 물을 수 있으며, 다양한 증명을 제시할 수 있었다.

증명 평가과제 1, 2에서 다룰 정리를 선정한 이후에는 각 과제에 포함할 증명들을 작성해야 했다. 증명 평가과제의 자료가 신뢰도와 타당도를 측정하는 데에 보다 신뢰할 만한 자료가 되기 위해서는 학생들이 증명 평가과제를 수행할 때에 실제로 증명 동료평가를 할 때와 비슷한 수행을 해야 한다. 따라서 학생들이 실제 증명 동료평가를 할 때와 최대한 비슷한 상황을 제공하기 위해, 실제로 학생들이 동료평가 시에 제출하였던 답안을 바탕으로 증명 평가과제의 평가 답안들을 구성하였다. 그러나 학생들의 답안을 그대로 제시하지는 않았다. 평가할 증명들의 다양성을 확보하기 위해서 연구자가 의도적으로 작성한 증명도 있으며, 다섯 개의 증명들이 받을 것으로 예상되는 논리성, 명료성, 참신성의 점수가 너무 치우치지 않도록 일부 답안을 수정하였다[5].

5) 연구에서 설계하고 사용한 '증명 평가과제 1'과 '증명 평가과제 2'의 내

3.2. 절차

본 연구에서는 증명 동료평가의 타당도를 조사하기 위하여 <정수론> 수강생들과 전문가들에게 각각 증명 5개를 평가하도록 요청하였다.

3.2.1. 증명 평가과제의 절차 - <정수론> 수강생

학생들은 학기 중반(8주차)과 학기가 끝난 이후에 각각 증명 평가과제 1과 증명 평가과제 2를 수행하였다. 연구자는 학생들에게 주어진 5가지의 증명이 동료가 작성한 증명이라고 가정한 상태에서 주어진 증명을 주별 동료평가에서 사용한 채점기준과 동일한 기준으로 평가할 것을 요청하였다. 또한 연구자는 학생들에게 해당 점수를 준 이유를 적도록 요청하였다. 증명 평가과제 1에는 총 20명의 수강생이 참여하였고 증명 평가과제 2에는 총 14명의 수강생이 참여하였다. '증명 평가과제 1'은 학생들이 소수의 무한성에 관한 증명을 배운지 얼마 지나지 않은 학기 중반에 이루어졌고, '증명 평가과제 2'는 학기가 종료된 뒤에 이루어졌다. 증명 평가과제 1은 오프라인에서 이루어졌으며 증명 평가과제 2는 온라인으로 진행되었다. 증명 평가과제 2는 학기가 끝난 이후에 이루어졌기 때문에 오프라인으로 자료를 수집하는 데에 어려움이 있어 온라인 설문조사의 형태로 자료를 수집하였다.

3.2.2. 증명 평가과제의 절차 - 전문가

전문가 2인에게도 증명 평가과제 1, 2를 수행하도록 하였다. 수강생들에게는 "제시된 증명이 동료가 작성한 것이라 가정하고, 각 증명을 현재 Classprep 과제에서 사용하고 있는 채점기준을 적용하여 채점해주세

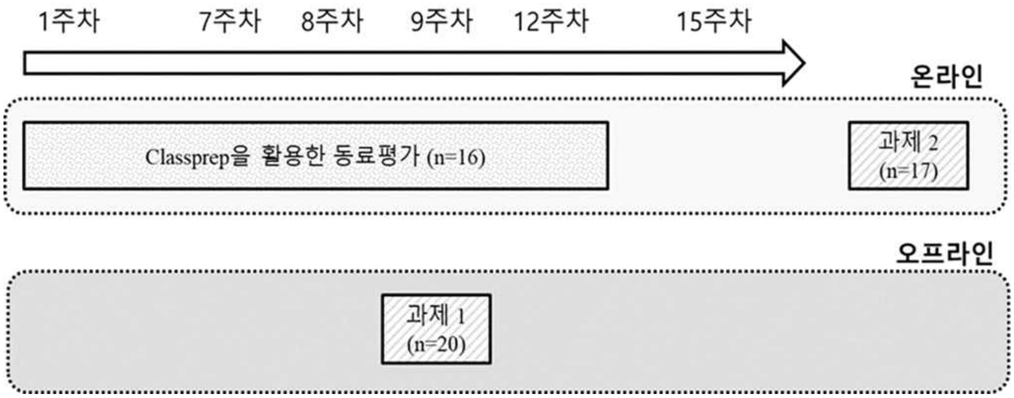
용은 부록에 수록하였다.

요. 그리고 그렇게 채점한 이유를 써주세요.” 라고 요청한 반면, 전문가들에게는 “제시된 증명을 학부생이 작성한 것이라 가정하고, 각 증명을 주어진 채점기준을 적용하여 채점하고, 그렇게 채점한 이유를 써주시길 바랍니다.” 라고 요청하였다. 참여한 전문가 1인은 대수학을 전공한 수학 박사였으며 다른 1인은 대수학을 전공하고 있는 박사과정 3년차 학생이었다. 이들은 본 연구의 학생들이 동료평가 시에 사용한 채점기준에 익숙하지 않았기 때문에 연구자는 과제 수행 전에 이들에게 채점기준에 대해 설명하는 시간을 가졌다. 이들의 답변은 전자우편을 통해 한글 문서파일의 형태로 수집되었다. 전문가들에게도 평가 시에 해당 점수를 준 이유도 함께 적도록 안내하였다.

4. 자료의 수집 및 분석방법

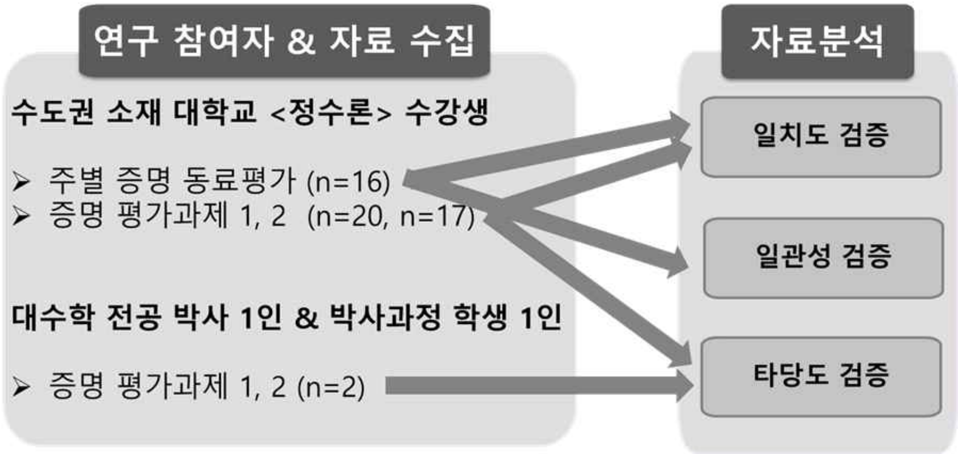
4.1. 자료 수집의 절차

본 연구의 전체적인 자료수집 절차는 [그림 Ⅲ-6]과 같다.



[그림 Ⅲ-6] 자료 수집의 절차

학생들은 1주차부터 12주차까지 7주차(중간고사 기간)를 제외한 11주에 매주 주별 온라인 동료평가 활동을 하였다[6]. 7주차와 15주차에 중간고사와 기말고사가 실시되었으며 중간고사 이후와 기말고사 이후에 1, 2차 증명평가과제를 실시하였다. [그림 Ⅲ-7]은 수집된 자료들이 각각 어떠한 분석에 사용되었는지를 도식화한 것이다.



[그림 Ⅲ-7] 각 분석에 사용된 자료들

주별 증명 동료평가 자료와 증명 평가과제 1, 2의 결과를 사용하여 일치도 검사를 실시하였으며, 주별 증명 동료평가 자료를 통해 일관성 검증을 실시하였고, 학생들이 작성한 증명 평가과제 1, 2 자료와 전문가가 작성한 증명 평가과제 1, 2 결과를 통해 타당도 검증을 실시하였다.

4.2. 증명 동료평가의 신뢰도 검증

본 연구에서는 해당 수업의 학생들의 증명 동료평가 결과를 바탕으로 대학생들의 증명 동료평가의 신뢰도를 살펴보았다. 어떠한 평가의 신뢰

6) 개별학생이 주별로 받은 동료평가 점수의 평균은 부록에 수록하였다.

도는 일치도(equivalence), 일관성(consistency) 그리고 안정성(stability)의 세 가지 측면에서 조사할 수 있다. 일치도란 동일한 평가를 서로 다른 평가자가 실시하였을 때에 얼마나 비슷한 결과가 나오는가에 대한 척도이다. 본 연구에서 다룬 동료평가에서 일치도는 3명의 학생이 동일한 증명을 채점하였을 때에 그 결과가 얼마나 비슷한가에 해당한다. 일관성은 일정 기간에 걸쳐 동일한 피평가자의 능력을 측정하였을 때에, 서로 다른 시점에 이루어진 평가 결과가 얼마나 유사한지를 의미한다. 본 연구에서 다룬 동료평가에서의 일관성은 한 학기에 걸쳐 이루어진 증명 동료평가에서 개별학생이 받은 점수가 얼마나 일관적인가를 의미한다. 이러한 일관성의 측면에서 신뢰도를 측정하는 경우는 측정되는 피평가자의 능력이 특정 기간에 걸쳐 일정하다는 가정을 한다는 점에 유의할 필요가 있다. 마지막으로 안정성은 같은 평가를 서로 다른 집단에 실시하였을 때에 얼마나 비슷한 결과가 나오는가에 관한 것이다. 해당 연구에서는 한 집단에만 동료평가를 실시하였으므로 안정성의 조사는 불가능하였다. 따라서 대학생들의 증명 동료평가의 신뢰도를 조사하기 위하여 안정성을 제외한 일치도와 일관성만을 측정하였다.

먼저, 한 학기 동안 수집된 증명 동료평가 자료를 통해 3명의 채점자 간 일치도를 측정하였다. 동일한 자료로 각 학생별로 얼마나 일관적인 평가 점수를 받았는지를 측정함으로써 증명 동료평가의 일관성을 측정하였다. 이와 더불어 추가로 수집한 ‘증명 평가과제 1’과 ‘증명 평가과제 2’의 결과 자료를 통하여 각각 20명과 14명의 채점자 간에 점수가 얼마나 일치하는 지를 검사하였다. 모든 일치도 검사와 일관성 검사에서는 크론바흐 알파 테스트를 사용하여 급내 상관계수를 구하였다. 급내 상관계수 값은 보통 0에서 1사이에 분포하며(Taylor, 2010) 추정 오차가 클 경우에는 음의 값이 측정되기도 한다. 자료들 간의 일치도가 높을수

록 급내 상관계수 값은 높다.

4.3. 증명 동료평가의 타당도 검증

동료평가의 타당도를 조사하는 방법 중 하나로 학생들이 얼마나 전문가와 유사한 평가를 하는지를 조사하는 방법이 있다. 따라서 학생들이 전문가와 유사한 평가를 하는지 조사함으로써 학부 수업에서 실시하는 증명 동료평가의 타당도를 조사할 수 있다. 본 연구에서는 증명 동료평가의 타당도를 조사하기 위해 <정수론> 수강생들과 전문가를 대상으로 실시한‘증명 평가과제 1’과 ‘증명 평가과제 2’의 결과를 활용하여 수강생 집단과 전문가 집단 간의 증명 평가과제 점수의 상관을 조사하였다. 타당도를 보다 여러 가지 각도에서 살펴보기 위해 전체 점수 뿐 아니라 논리성, 명료성 그리고 참신성 점수들의 전문가 집단-학생 집단 상관도 조사하였다. 상관 계수를 구할 때에는 평가의 신뢰도 및 타당도 검사에서 널리 활용되는 크론바흐 알파 테스트를 사용하였다.

Ⅳ. 연구결과

학생들이 주별로 실시한 증명 동료평가의 결과를 토대로 증명 동료평가의 신뢰도를 일치도와 일관성 측면에서 검토하였다. 또한 추가적으로 실시한 증명 평가과제의 결과를 토대로 증명 동료평가의 타당도와 신뢰도를 검토하였다. 이 장에서는 증명 동료평가의 신뢰도와 타당도 조사 결과를 살펴볼 것이다.

1. 신뢰도

평가의 신뢰도를 검증하는 방법으로 대표적으로 세 가지를 꼽을 수 있다. 일치도, 일관성 그리고 안정성이 그에 해당한다. 안정성은 동일한 평가를 서로 다른 집단에 적용하였을 때에 평가 결과가 얼마나 유사한지를 의미한다. 본 연구에서는 한 집단에 적용한 동료평가 자료만을 이용하므로 신뢰도 검증에서 안정성 검사는 제외하였다. 이에 따라 이번 절에서는 증명 동료평가의 일치도와 일관성을 검토할 것이다.

1.1. 일치도

일치도는 같은 답안을 서로 다른 평가자가 평가하였을 때에 얼마나 비슷한 결과가 나오는 지를 의미한다. 평가자들 간의 평가결과가 유사할수록 일치도가 높은 평가라고 할 수 있다. <정수론> 수강생들이 주별로 실시한 증명 동료평가는 하나의 답안 당 평가자가 3명씩 배정되었다. 따라서 증명 동료평가 자료를 통해 일치도 검사를 할 때에는 표본의 크기가 3인 자료들을 사용하게 되었다. 더불어 증명 평가과제에서 수집된 학생들의 평가 결과를 사용하여 일치도 검사를 다시 한 번 실시하였다. 일치

도 검사 시에는 각 채점자를 하나의 항목으로 취급하여 급내상관계수를 측정하였으며 이때 급내상관계수를 측정하기 위해 크론바흐 알파 방법을 사용하였다.

1.1.1. 주별 동료평가에서의 채점 일치도

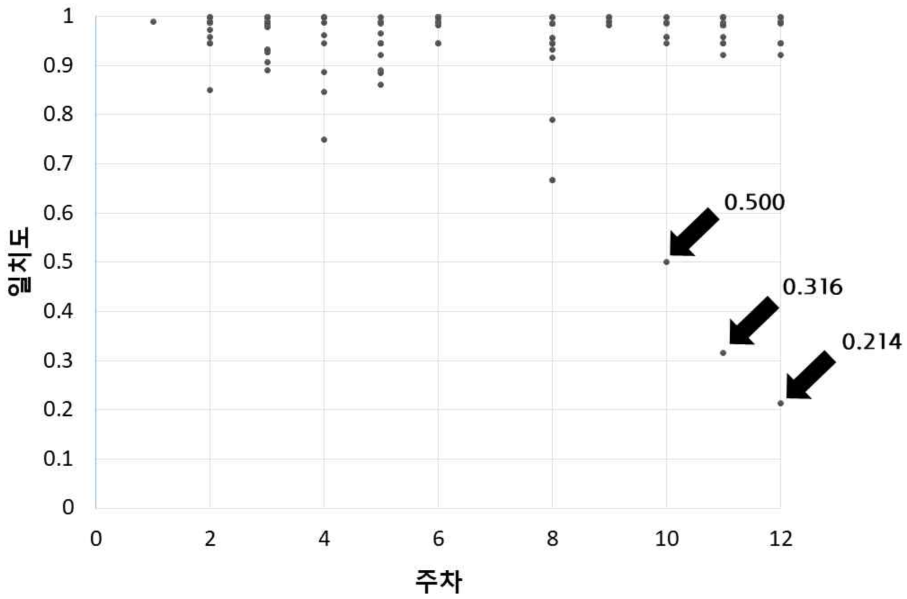
<정수론> 수강생 25명 중 자신의 동료평가 활동 자료를 연구에 활용하는 것에 동의한 학생은 총 16명이었다. 이에 따라 연구자는 25명 학생 간에 이루어진 증명 동료평가 자료 중에서 16명의 학생들끼리 동료평가가 이루어진 자료만을 분석하였다. 즉, 자신의 동료평가 활동 자료의 사용을 동의하지 않은 9명의 학생이 평가한 결과나 평가받은 결과는 분석 시에 제외하였다. 9명의 학생의 자료를 제외하고 난 뒤 16명의 학생들이 받은 평가 중에 3명의 평가 결과가 모두 남은 경우만을 가지고 일치도 조사를 실시하였다. 예를 들어 학생 A가 학생 B, 학생 C, 학생 D로부터 평가를 받았는데 학생 B가 자료의 활용에 동의하지 않았다면 학생 A가 받은 점수는 일치도 조사에서 제외하였다. 동료평가를 수행하지 않은 학생으로 인해 자료에 결손이 생겨, 한 학생이 2인 이하의 평가자로부터 평가를 받은 경우도 분석에서 제외되었다. 이렇게 조건에 맞는 자료들만 추려내자 과제별로 많지 않은 학생들이 받은 점수만이 남게 되었다. 그렇게 남은 점수들을 대상으로 알파 분석을 통해 급내 상관 계수를 측정하는 방법으로 일치도 분석을 진행하였다[7]. 다음 <표 Ⅳ-1>은 총 143개의 표본에 대해 채점자간 일치도를 측정한 결과를 정리한 것이다.

7) 총 143회의 일치도 검사의 결과는 부록에 수록하였다.

<표 Ⅳ-1> 주별 동료평가의 채점 일치도

크론바흐 알파 계수	횟수	신뢰도
0.90 이상	129회 (90.2%)	신뢰도 아주 높음 (Very highly reliable)
0.80 이상 0.90 미만	7회 (4.9%)	신뢰도 높음 (highly reliable)
0.70 이상 0.80 미만	2회 (1.4%)	신뢰할 만 함 (reliable)
0.60 이상 0.70 미만	2회 (1.4%)	미미하게 신뢰할 만 함 (marginally/minimally reliable)
0.60 미만	3회 (2.1%)	용납할 수 없게 신뢰도 낮음 (unacceptably low reliability)

0.6 미만의 값이 3회(0.214, 0.316, 0.500), 0.6 이상 0.7 미만의 값이 2회(0.667, 0.667), 0.7 이상 0.8 미만이 2회(0.750, 0.789) 0.8 이상 0.9 미만이 7회(0.846, 0.851, 0.861, 0.885, 0.889, 0.892, 0.892) 나온 것을 제외한 129회의 검사에서 0.9 이상의 값이 도출되었다. 이러한 결과는 하나의 증명에 대한 세 명의 평가자들이 준 점수들이 서로 매우 유사함을 의미한다. 다음 [그림 Ⅳ-1]은 주차별 채점 일치도의 표본을 그래프로 나타낸 것이다.



[그림 Ⅳ-1] 주차별 채점 일치도 분포

위의 그림을 통해 일치도가 아주 낮은 표본들(0.500, 0.316, 0.214)이 10주차, 11주차, 12주차에 분포하고 있음을 확인할 수 있다. 학기말에 이르러서 학생들이 보다 신뢰도 있는 증명을 하게 되었을 것 같은데, 학기초에는 나타나지 않았던 아주 낮은 신뢰도 계수가 학기말에 나타난 것이다. 이러한 현상이 일어난 것은 학기 후반에 이르면서 다루는 수학적 내용이 어렵고 복잡해졌고, 이에 따라 한 정리에 대한 다양한 증명이 가능해졌으며 평가자가 자신이 평가하는 증명의 내용을 제대로 이해하지 못하는 경우가 생겼기 때문으로 보인다. 해당 결과를 통해 다른 영역의 평가에 비해 논리성의 평가 결과가 큰 차이가 난다는 것을 확인할 수 있다. 이에 따라 이어지는 절에서는 채점자간 일치도가 아주 낮게 나온 세 표본에서 학생들의 평가 내용이 어떠한지를 살펴보도록 한다.

1.1.2. 0.6미만의 일치도를 나타낸 표본의 내용 분석

이 절에서는 채점자간 일치도가 0.6미만으로 아주 낮게 나온 세 표본에서 학생들의 평가 내용이 어떠한지를 살펴보도록 한다. 명료성과 참신성의 평가 내용보다 논리성 영역의 평가 내용에서 큰 차이가 나타나기 때문에 논리성 영역의 평가를 위주로 살펴본다. 다음 <표 Ⅳ-2>는 학생C가 정리 4.36의 논증에 대해 받은 동료평가 결과 점수를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-2> 학생 C가 작성한 정리 4.36의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	3	1	1	0.500
명료성	2	2	3	
참신성	1	1	1	

한 학생은 논리성 영역에 3점을 부여한 반면 나머지 두 학생은 논리성 영역에 1점을 부여하였다. 이 학생들이 작성한 논리성 영역 평가 근거를 살펴보면, 평가자 학생 3인 모두 학생 C가 작성한 논증에서 존재성(existence)은 보이지 않고 유일성(uniqueness)만 보인 것이 논리적인 결함이라고 보았다. 이들은 이처럼 동일한 부분을 지적하면서도 논리성 점수를 부여함에 있어서는 다소 차이를 보였다. 즉, 같은 결점을 발견하고도 어떤 평가자는 논리적으로 큰 결함이라고 생각한 반면 어떤 평가자는 그리 크지 않은 결함이라고 생각한 것이다. 한 증명에 논리적인 결함이 있을 때에 어떤 평가자는 그 결함이 중대하다고 여길 수도 있는 한편 다

른 평가자는 그 결함이 사소하다고 여길 수도 있다. 이러한 이유에서 논리성에 결함이 있는 증명의 경우에는 평가에서 평가자들의 주관적 판단이 관여되기 때문에 논리성에 결함이 없는 증명에 비해 채점자간 일치도가 낮은 현상이 나타날 것으로 예상할 수 있다.

한편 학생 M이 정리 6.4의 논증에 대해 받은 평가와 학생 K가 정리 6.23의 논증에 대해 받은 평가는 조금 다른 양상을 보인다. 다음 <표Ⅳ-3>은 학생 M이 11주차에 정리 6.4의 논증에 대해 받은 동료평가 결과 점수를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-3> 학생 M이 작성한 정리 6.4의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	5	1	1	0.316
명료성	3	3	3	
참신성	2	1	1	

<표 Ⅳ-3>에서 논리성 영역의 평가를 보면, 한 학생은 논리성 영역에 만점을 주었으나 나머지 두 학생은 논리성 영역에 1점을 주었음을 확인할 수 있다. 이 학생들이 작성한 피드백 내용을 살펴보면 평가자 1은 “논리적인 증명인 것 같다”고 적었으나 평가자 2와 평가자 3은 해당 증명에서 사용한 아이디어가 잘못되었음을 지적하였다. 실제로 학생 M이 작성한 증명에서는 평가자 2와 평가자 3이 지적한 것처럼 논리적으로 중대한 결함이 있었다. 즉 세 명의 평가자 중 2인은 해당 증명의 논리적인 결함을 발견하고 논리성 영역에서 1점을 부여하였으나 나머지 1인은

해당 증명의 논리적인 결함을 발견하지 못한 채 논리성 영역에 만점인 5점을 부여하였다. 이는 학생들의 증명 검증 능력의 차이가 낮은 채점 일치도를 야기하였다고 해석할 수 있다. 다음으로 살펴볼 표본에는 위 두 가지 표본에서 살펴본 현상이 모두 나타나 있다. 다음 <표 Ⅳ-4>는 학생 K가 12주차에 정리 6.23의 논증에 대해 받은 동료평가 결과 점수를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-4> 학생 K가 작성한 정리 6.23의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	1	5	3	0.214
명료성	3	2	3	
참신성	1	1	2	

<표 Ⅳ-4>에서 학생 K가 작성한 증명의 논리성에 대해 세 학생이 부여한 점수가 모두 다름을 확인할 수 있다. 논리성 영역에서 평가자 2는 만점인 5점을 부여한 한편 평가자 3은 3점을, 평가자 1은 1점을 부여하였다. 평가자 학생들이 작성한 피드백 내용을 살펴보면 평가자 2는 “논리적 흐름으로는 이해가 가는 증명”이라고 언급하였고 평가자 3은 “phi(n)개의 집합을 어떻게 구성하는 것인지 잘 이해가 안 된다. 조금 더 자세한 설명이 필요하다”고 지적하였다. 반면 평가자 3은 “phi(m)과 phi(n)에 대해 언급한 내용의 위치가 바뀌어야 한다”고 지적하며 논리성 영역에 1점을 부여하였다. 평가자 1과 평가자 3은 해당 논증의 결함을 발견하였으나 평가자 2는 아무런 결함을 발견하지 못했다. 또한 평

가자 1은 해당 결함이 중대하다고 여기고 논리성에 1점을 부여한 한편 평가자 3은 해당 결함이 중대한 것은 아니라고 보고 논리성에 3점을 부여하였다. 이 표본에서는 학생들의 검증 능력의 차이가 평가자 2와 나머지 평가자 2인의 평가 간의 차이를 야기했으며, 해당 논증의 논리적인 결함이 중대한지에 대한 의견 차이가 평가자 1과 평가자 3의 평가 간의 차이를 야기했다.

신뢰도가 낮게 나온 표본들을 살펴본 결과, 평가자들의 증명 검증 능력의 차이가 낮은 신뢰도의 원인이 되기도 했고 논리적인 결함의 중대성에 대한 판단이 달라 낮은 신뢰도가 나타나기도 했다. 또한 살펴본 세 표본에서 다룬 논증 모두 논리적인 결함이 있었음에 주목할 만하다. 평가하는 논증에 논리적인 결함이 없다면 일부 평가자는 논리적인 결함을 발견하고 다른 평가자는 발견하지 못하는 현상이 발생하지 않을 것이다. 또한 논리적인 결함이 없다면 발견된 결함이 중대한지를 판단하는 과정이 필요하지 않다. 따라서 논리적인 결함이 없는 논증은 낮은 일치도를 보이기 어렵고 논리적인 결함이 있는 논증은 낮은 일치도를 보일수도 있다고 할 수 있다.

이상에서는 신뢰도가 낮게 나온 세 개의 표본을 구체적으로 살펴보고 낮은 신뢰도가 무엇에서 기인하였는지를 살펴보았다. 이하에서는 일치도가 1이 측정된 표본 45개와 0.9 이상 1 미만의 일치도를 나타낸 표본 85개를 살펴보고 해당 표본들에 어떠한 특징이 있는지 살펴본다.

1.1.3. 0.6이상 0.7미만의 일치도를 나타낸 표본의 내용 분석

이 절에서는 채점자간 일치도가 0.6이상 0.7미만이 나온 두 개의 표본에서 학생들의 평가 내용이 어떠한지를 살펴보도록 한다. 다음 <표 Ⅳ

-5>는 학생 H가 작성한 정리 3.13의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-5> 학생 H가 작성한 정리 3.13의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	2	3	3	0.667
명료성	2	3	3	
참신성	2	2	1	

해당 표본의 경우에는 세 명의 평가자가 각 항목에 부여한 점수가 모두 달랐다. 다시 말해 같은 평가를 한 평가자들이 없었다. 또한 평가자 1은 논리성에 2점을 부여하였는데 학생들에게 주어진 채점 기준에 따르면 논리성에 부여할 수 있는 점수는 5점, 3점 그리고 1점 뿐이다. 입력상의 오류인지 의도한 것인지 알 수 없다. 이들이 작성한 피드백 내용을 살펴보면 평가자 1과 평가자 3은 논리성 영역의 피드백 내용을 길게 작성하였는데 이들이 작성한 피드백 내용이 많은 부분 일치함을 발견할 수 있다. 그러나 한 학생은 논리성에 2점을 다른 한 학생은 3점을 부여하였다. 본 표본에서도 같은 근거를 제시하면서도 서로 다른 점수를 부여하는 현상이 나타났다. 또한 평가자 1과 평가자 2 모두 본 논증이 참신하다고 평가하였는데 평가자 1은 “정리 2.46을 사용하려는 시도가 좋았다”고 그 근거를 들었고 평가자 2는 “나름대로 참신한 풀이다”라고 근거를 작성하였다. 이처럼 같은 점수를 부여하면서도 한 학생은 참신하다고 생각하는 근거를 구체적으로 작성하였으나 다른 학생은 주관적인

근거를 제시하였다. 계속해서 일치도가 0.6이상 0.7미만이 나온 표본의 평가 결과를 살펴본다. 다음 〈표 Ⅳ-6〉은 학생 J가 작성한 정리 3.28의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

〈표 Ⅳ-6〉 학생 J가 작성한 정리 3.28의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	1	3	5	
명료성	3	3	3	0.667
참신성	1	1	1	

해당 표본의 경우 명료성과 참신성 영역에서 세 명의 평가자가 부여한 점수는 일치하는 반면 논리성 영역의 점수는 다양한 분포를 보인다. 학생들이 작성한 논리성 영역의 피드백 내용을 살펴보면 평가자 3은 본 논증이 “논리 그 자체이다.” 라고 작성한 반면 학생 G는 “귀납법은 무한히 많은 자연수 n에 대해 명제가 성립함을 보일 때에 쓰는 것인데 본 증명에서는 유한개의 자연수에 대해서만 명제가 성립함을 보이면 된다. 왜 굳이 수학적 귀납법을 썼는지 모르겠다.” 라고 피드백을 작성하였다. 학생 O는 “수학적 귀납법에 의해 마지막 부분을 증명하면 모든 자연수에 대해 주어진 성질을 만족한다고 작성한 것이므로 이 증명은 옳지 않다” 고 피드백을 작성하였다. 이 표본에서도 일부 평가자는 논증의 논리적인 결함을 발견하는 한편 다른 평가자는 해당 결함을 발견하지 못함으로 인해 평가 결과의 일치도가 낮아지는 현상이 나타났다. 또한 같은 결함을 발견하고도 해당 결함이 중대하다고 여기는지의 여부에 따라 평가

결과가 달라지는 현상을 나타내기도 했다. 이러한 결과는 1.1.2.절에서 살펴본 결과와 일치하는 결과이다.

1.1.4. 0.7이상 0.8미만의 일치도를 나타낸 표본의 내용 분석

이 절에서는 채점자간 일치도가 0.7이상 0.8미만이 나온 두 개의 표본에서 학생들의 평가 내용이 어떠한지를 살펴보도록 한다. 다음 <표 Ⅳ-7>은 학생 P가 작성한 정리 2.7의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-7> 학생 P가 작성한 정리 2.7의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	5	1	5	0.750
명료성	3	1	3	
참신성	1	1	1	

해당 표본의 결과를 살펴보면 평가자 2가 부여한 점수가 평가자 1과 평가자 3이 부여한 점수와 큰 차이를 보임을 알 수 있다. 평가자 2는 피드백에서 “이 증명은 정리 2.1의 증명인 것 같다. 이 증명 자체에는 논리적인 결함이 없으나 다른 정리의 증명을 올렸다. 혹시 제가 잘못 생각했다면 피드백 주세요.” 고 서술하였다. 실제로 학생 P가 작성한 논증에서는 정리 2.7에서 서술한 것을 보이고 있었다. 즉 평가자 2가 해당 논증을 잘 이해하지 못한 것으로 인해 다른 평가자들과 다른 평가를 하게 되었고 이러한 점이 낮은 신뢰도를 야기하였다. 계속해서 일치도가 0.7

이상 0.8미만이 나온 표본의 평가 결과를 살펴본다. 다음 <표 Ⅳ-8>은 학생 A가 작성한 정리 3.13의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-8> 학생 A가 작성한 정리 3.13의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	3	5	3	0.789
명료성	3	3	3	
참신성	2	2	2	

해당 표본의 결과를 살펴보면 세 명의 평가자가 명료성과 참신성에 부여한 점수는 동일하지만 논리성 영역에 부여한 점수에 차이가 있다. 이들이 논리서 영역에 작성한 피드백을 살펴보면 평가자 2는 간략하게 “논리적인 증명이다” 라고 서술하였다. 한편 평가자 1과 평가자 3은 해당 논증에서 논리적인 결함으로 보이는 부분을 지적하였다. 그러나 두 평가자가 지적한 부분은 서로 달랐다. 해당 표본의 결과를 통해서 평가자마다 논리적으로 설명이 불충분하다고 느끼는 지점이 다를 수 있음을 확인할 수 있다. 또한 서로 다른 지점을 지적하면서 같은 점수를 부여할 수 있음을 보여준다.

1.1.5. 0.8이상 0.9미만의 일치도를 나타낸 표본의 내용 분석

이 절에서는 채점자간 일치도가 0.8이상 0.9미만이 나온 7개의 표본 중 주차가 겹치는 것들을 제외한 4개의 표본에서 학생들의 평가 내용이 어떠한지를 살펴보도록 한다.8) 다음 <표 Ⅳ-9>는 학생 O가 작성한 정리

2.7의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-9> 학생 O가 작성한 정리 2.7의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	3	5	3	0.846
명료성	3	3	3	
참신성	2	1	1	

세 명의 평가자가 부여한 점수를 살펴보면 명료성의 평가는 일치하지만 논리성과 참신성의 평가는 차이를 보인다. 평가자 2는 전체적으로 피드백이나 평가 근거를 남기지 않았다. 평가자 1은 논리성 영역에서 “증명이 잘 완성 된 것 같다” 라고 피드백을 남겼는데 논리성 점수는 3점을 부여하였다. 평가자 3은 논리성 영역에서 “논리적인 증명이었다. 그러나 사소한 부분에서 조금 주의할 부분이 있다.” 라고 피드백을 시작하며 보다 꼼꼼한 증명이 되기 위해서 유의해야할 부분을 제언하였다. 평가자들이 서로 다른 피드백을 하면서도 같은 점수를 부여하는 현상이 나타났다. 참신성 평가의 피드백에서 평가자 1은 “증명 방법이 참신하다” 고 서술하였으며 평가자 3은 “정석적인 풀이였다” 고 서술하였다. 두 학생 모두 주관적인 근거를 제시하였다. 또한 해당 표본에서 논리적이라고 평가하는 경우, 명료하다고 평가하는 경우, 참신하지 않다고 평가하지 않는 경우 피드백이 간략한 현상이 나타났다. 계속해서 일치도가 0.8이상 0.9미만이 나온 표본의 평가 결과를 살펴본다. 다음 <표 Ⅳ-10>은 학생

8) 7개 표본의 평가 결과는 부록에 수록하였다.

K가 작성한 정리 1.22의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

〈표 Ⅳ-10〉 학생 K가 작성한 정리 1.22의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	3	5	3	0.851
명료성	3	3	2	
참신성	2	1	1	

본 표본에서는 세 가지 영역 모두에서 평가자들 간의 점수 차이가 나타났다. 이들이 작성한 피드백을 살펴보면 논리성 영역의 평가에서 평가자 1과 평가자 2는 같은 부분을 지적하며 정당화가 필요하다고 서술하였다. 반면 평가자 2는 “논리적인 증명”이라고 간략하게 서술하였다. 해당 표본에서도 일부 평가자는 논증의 논리적인 결함을 발견하는 반면 다른 평가자는 발견하지 못한 것으로 인해 채점자간 일치도가 낮아졌다. 또한 해당 표본에서 논리적이라고 평가하는 경우, 명료하다고 평가하는 경우, 참신하지 않다고 평가하지 않는 경우 피드백이 간략한 현상이 나타났다. 계속해서 일치도가 0.8이상 0.9미만이 나온 표본의 평가 결과를 살펴본다. 다음 〈표 Ⅳ-11〉은 학생 M이 작성한 정리 2.34의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-11> 학생 M이 작성한 정리 2.34의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	3	5	5	0.885
명료성	3	2	3	
참신성	1	2	1	

해당 표본의 경우에도 모든 영역의 평가에서 평가자들 간의 점수 차이가 나타났다. 평가자 1의 경우에는 “해당 논증에서 k가 1인 경우와 1보다 큰 경우를 나누어서 다루지 않았으나 본 정리의 증명에서 경우를 나누어 증명하는 것이 매우 중요하다” 고 지적하였다. 반면 평가자 2와 평가자 3은 특별한 지적 없이 논리성에 만점을 부여하였다. 참신성 평가에서 평가자 2는 “정리 2.1을 사용한 것이 참신하다” 고 서술하였고 나머지 2인은 참신성에 대한 피드백을 남기지 않았다. 해당 표본에서 논리적이라고 평가하는 경우, 명료하다고 평가하는 경우, 참신하지 않다고 평가하지 않는 경우 피드백이 간략한 현상이 나타났다. 계속해서 일치도가 0.8이상 0.9미만이 나온 표본의 평가 결과를 살펴본다. 다음 <표 Ⅳ-12>는 학생 C가 작성한 정리 1.34의 논증에 대해 받은 동료평가 결과를 정리한 것이다. 평가 시에 평가자 학생들이 제시한 피드백은 부록에 수록하였다.

<표 Ⅳ-12> 학생 C가 작성한 정리 1.34의 논증 동료평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	5	5	3	0.892
명료성	3	3	3	
참신성	2	2	2	

해당 표본의 경우 명료성과 참신성의 경우 세 평가자가 부여한 점수가 일치했다. 논리성 평가에서 평가자 1은 “아주 논리적이다”라고 간략한 근거를 들며 5점을 부여하였고, 평가자 2는 “처음 집합 S를 정의할 때 x가 빠진 것 빼고는 논리적인 결함이 없다”라고 서술하며 5점을 부여하였다. 평가자 3의 경우에는 논리성 피드백에서 “S를 정의할 때에 x, y가 Z의 원소임을 명시해주어야 하는 것 아닌가요?”라고 질문하였다. 해당 표본에서도 일부 평가자는 논리적인 결함이라고 여기지 않는 부분을 다른 평가자는 논리적인 결함으로 여길 수 있는 현상으로 인해 일치도가 낮아졌다.

1.1.6. 0.9이상의 채점 일치도를 나타낸 표본의 내용 분석

이 절에서는 일치도가 1이 측정된 표본 44개와 0.9 이상 1 미만의 일치도를 나타낸 표본 85개를 살펴보고 해당 표본들에 어떠한 특징이 있는지 살펴본다. <표 Ⅳ-13>은 일치도 1이 나온 표본 44개의 평가 결과를 나타낸 것이다.

<표 Ⅳ-13> 일치도 1이 측정된 표본 44개의 평가 결과

	평가자 1	평가자 2	평가자 3	일치도
논리성	5	5	5	1
명료성	3	3	3	
참신성	1	1	1	

일치도가 1이 측정된 표본 44개 모두에서 동일한 점수 분포가 나타났다. 44개의 표본에서 평가자 세 명 모두가 논리성에 5점, 명료성에 3점 그리고 참신성에는 1점을 부여하였다.[9] 일치도가 1인 표본들 중에 채점자 3인 모두가 논리성에 5점, 명료성에 3점, 그리고 참신성에 1점을 부여한 표본이 압도적으로 많은 현상은 앞서 1.1.2절에서 일치도가 낮은 표본들을 분석하여 도출해낸 결과와 연결해서 살펴보면 설명이 가능하다. 일치도가 낮은 표본들의 경우 논증에 논리적인 결함이 있었다. 논리적인 결함이 있는 경우 평가자들의 검증 능력 차이나 주관적인 판단으로 인해 평가의 차이가 나타났다. 반면 논리적인 결함이 없는 증명의 경우에는 일부 평가자는 논리적인 결함을 발견하고 나머지 평가자는 발견하지 못하는 현상이 나타나지 않으며, 평가자들이 발견된 논리적인 결함의 중대성을 판단하는 과정이 필요하지 않다. 따라서 논리적인 결함이 없는 증명의 경우에는 세 명의 평가자 모두 논리성에 5점을 부여할 것이다. 비슷하게 명료하게 서술되어 있는 논증의 경우, 명료성을 평가함에 있어서 학생들의 주관적인 판단이 개입될 여지가 적을 것이다. 그렇기 때문에 명료한 논증의 경우 평가자 모두 명료성에 3점을 부여할 것으로 예상된다. 한편 참신성 평가에서는 참신하지 않은 논증일수록 일치도가 높을

9) 일치도가 1이 측정된 표본들 중 일부의 평가 내용을 부록에 수록하였다.

것으로 보인다. 왜냐하면 참신한 아이디어를 담고 있는 논증의 경우, 일부 평가자는 참신하다고 판단하는 한편 다른 평가자는 참신하지 않다고 판단할 수 있다. 반면 일반적인 아이디어를 담고 있는 논증의 경우에는 대부분의 평가자들이 참신하지 않다고 판단할 것이다. 이러한 이유에서 참신하지 않은 증명일수록 참신성 평가의 일치도가 높을 것이라고 예상할 수 있다.

0.9 이상 1 미만의 일치도를 보인 표본 총 85개를 살펴본 결과, 이들 중 53개의 표본의 경우에 각 논증을 평가한 3명 중 2명의 평가가 논리성, 명료성, 그리고 참신성을 각각 5점, 3점 그리고 1점을 부여하였다. 이를 종합하면, 높은 일치도를 보인 표본들에서 논리성에 5점, 명료성에 3점 그리고 참신성에 1점이 자주 부여되는 현상이 나타났다.

본 절에서는 학생들이 주별로 실시한 동료평가의 채점자간 일치도를 살펴보았다. 분석 결과 주별 증명 동료평가의 채점자간 일치도는 대부분의 경우에 매우 높았으며 평가 과정에서 주관적인 판단이 수반되거나 평가자들 간에 배경지식이나 검증 능력의 차이가 있을 때에 채점자간 일치도가 낮을 수도 있음을 확인하였다. 일치도가 높은 표본들의 경우에는 해당 평가에서 평가자의 주관적인 판단이 수반되지 않았다는 특징이 있었다. 한편, 본 일치도 조사에서 사용된 표본들의 크기가 3으로 매우 작다. 연구자는 이처럼 작은 표본의 크기가 채점자간 일치도 측정에 영향을 주었을 수도 있다고 판단하여 보다 크기가 큰 표본에서 일치도 검사를 실시하였다. 이어지는 절에서는 주별 동료평가 이외에 따로 수집된 자료인 '증명 평가과제 1'과 '증명 평가과제 2' 결과 자료를 분석하여 각각 크기 20과 17의 표본의 채점자간 일치도를 살펴본다.

1.1.7. 증명 평가과제에서의 채점 일치도

주별 동료평가 과제를 토대로 채점가간 일치도를 검사했을 때에 일치도가 대부분 0.9 이상으로 매우 높게 나왔으나 수집된 표본들의 크기가 모두 3으로 매우 작았다. 연구자는 이처럼 작은 표본의 크기가 채점자간 일치도 측정에 영향을 주었을 수도 있다고 판단하여 보다 크기가 큰 표본에서 일치도 검사를 실시하였다. 수집된 ‘증명 평가과제 1’과 ‘증명 평가과제 2’의 결과를 통해 표본의 크기가 각각 20과 17일 때의 채점자간 일치도를 조사하였다. 주별 동료평가 자료의 일치도 검사 시와 동일하게 크론바흐 알파 분석을 통해 급내 상관 계수를 측정하였다. ‘증명 평가과제 1’과 ‘증명 평가과제 2’에서 학생 집단의 채점자간 일치도를 조사한 결과는 다음 <표 Ⅳ-14>와 같다.

<표 Ⅳ-14> 증명 동료평가에서 학생 집단의 채점자간 일치도

	전체 점수	논리성	명료성	참신성
증명 평가과제 1	0.943	0.963	0.943	0.915
증명 평가과제 2	0.970	0.727	0.854	0.915

채점자간 일치도 조사 결과, ‘증명 평가과제 1’에서는 전체 점수 뿐 아니라 논리성, 명료성 그리고 참신성 점수 모두 매우 높은 채점자간 일치도를 보였다. 학생들이 학기 중반에도 이미 서로 비슷한 채점을 하고 있었다고 볼 수 있다. ‘증명 평가과제 2’에서는 전체점수, 논리성 점수, 명료성 점수, 참신성 점수 모두 채점자간 일치도가 다소 높게 나왔

다. ‘증명 평가과제 1’에서는 모든 영역의 채점자간 일치도가 매우 높았던 것에 비해, ‘증명 평가과제 2’에서는 전체 점수의 채점자간 일치도는 매우 높았으나 각 영역 점수의 채점자간 일치도는 이에 비해 상대적으로 낮았다. 같은 증명을 평가할 때에 어떤 채점자는 논리성에 결함이 있다고 보고 논리성 점수를 낮게 주고 나머지 점수는 만점을 주고, 다른 채점자는 명료성에만 결함이 있다고 보고 명료성 점수는 낮게 주고 나머지 점수는 만점을 줄 수가 있다. 이때, 이 학생들이 준 영역별 점수는 차이가 있지만 이들이 부여한 총점은 같을 수 있다. 이러한 경우들로 인해서 ‘증명 평가과제 2’에서 전체 점수의 일치도가 각 영역별 점수의 일치도에 비해 더 높게 나온 것으로 보인다.

또한 ‘증명 평가과제 1’과 ‘증명 평가과제 2’의 결과를 비교해보면, 시간이 지나면서 전체 점수의 일치도는 소폭 상승했으나 논리성 점수와 명료성 점수의 일치도는 약간 감소했다. 시간이 지나면서 학생들간의 채점 일치도가 증가할 것으로 예상한 것과 다른 결과였다. 그러나 ‘증명 평가과제 1’에서의 일치도가 워낙에 높았으며, ‘증명 평가과제 2’에서의 일치도도 높은 수치라는 점에서 이러한 감소는 크게 시사하는 바가 없다고 보여 진다.

1.2. 일관성

일관성은 같은 피평가자의 능력에 대한 서로 다른 평가가 얼마나 일관성 있는가를 의미한다. 서로 다른 시점에 이루어진 평가들의 평가 결과가 유사할수록 일관성이 높은 평가라고 할 수 있다. 연구자는 학생들이 주별로 받은 동료평가 점수의 일관성을 조사하기 위하여 16명의 학생들이 매주 받은 네 개의 동료평가 점수의 평균을 구하였다.[10] 학생들은 주

---

10) 각 학생이 주별로 받은 동료평가 점수의 평균은 부록에 수록하였다. 전체

별로 4개의 증명 과제를 Classprep 상에 제출했으며 증명 당 3명의 평가자에 의해 채점된 결과를 받았다. 연구자는 학생들이 받은 12개의 평가 점수의 평균을 구했다. 제출하지 않은 과제가 있거나 채점자가 3인 미만인 과제가 있는 경우에는 해당 점수들을 결측값으로 취급하고 나머지 점수들의 평균을 구하였다. 또한 항목별로 학생들이 주별로 받은 점수의 평균을 구하였다. 그런 다음 각 학생이 한 학기 동안 받은 전체 점수와 항목별 점수의 일관성을 조사하였다. 일관성 조사에서는 점수들의 분산이나 상관계수 등이 주로 사용되는데 본 연구에서는 이때에 크론바흐 알파 테스트 사용해 상관계수를 측정하였다. 그 결과는 다음 <표 Ⅳ-15>와 같다.

<표 Ⅳ-15> 증명 동료평가 점수의 내적 일관성 계수

	전체 점수	논리성	명료성	참신성
일관성 계수	0.620	0.574	0.367	0.258

학생들이 한 학기 동안 받은 증명 동료평가 점수의 내적 일관성을 조사한 결과 그 일관성은 전체적으로 그리 높지 않았으며 특히 명료성 점수와 참신성 점수의 일관성은 낮은 편이었다. 이 결과는 심리학 학부 전공 수업에서 한 학기에 걸쳐 실시한 글쓰기 동료평가 점수의 내적 일관성이 높게 나온 배수정과 박주용(2016)의 연구 결과와 상반된다. 어떠한 평가의 낮은 일관성은, 평가 자체의 낮은 신뢰도 때문일 수도 있으나 다른 한편으로 평가되는 개인의 능력이 일정하지 않은 것 때문일 수도 있

점수뿐만 아니라 논리성 점수, 명료성 점수 그리고 참신성 점수의 주별 평균도 수록하였다. 논리성 점수의 만점은 5점, 명료성 점수의 만점은 3점, 참신성 점수의 만점은 2점이므로 전체 점수의 만점은 10점이다.

다. 따라서 낮은 일관성이 낮은 신뢰도를 함의한다고 단정할 수는 없다.

어떠한 평가의 일관성 검증은 여러 시기에 걸친 측정이 동일한 개인의 동일한 능력을 일관성 있게 측정해내는가를 검증하는 것이다. 즉, 측정되는 개인의 능력이 일관적이라는 가정이 있을 때에 일관성 검사가 의미를 가진다. 그러나 본 연구에 참여한 학생들은 대부분 학부 2학년 학생들로 엄밀한 증명의 작성을 배우는 시작 단계에 있는 학생들이었다. 따라서 한 학기 동안 증명에 대해 학습한 연구 참여자들의 증명 능력이 한 학기 동안 일정할 것이라고 가정하는 것은 다소 무리가 있다. 다시 말해, 연구 참여자 학생들은 한 학기 동안 증명에 대한 학습을 하였으므로 이들의 증명 작성 능력에 변화가 있었고, 이러한 증명 작성 능력의 변화는 증명 동료평가의 낮은 일관성을 야기했을 것으로 보인다.

배수정과 박주용(2016)의 연구에서는 일관성 검증을 위해 측정한 크론바흐 알파 계수가 전체 점수와 각 영역별 점수 모두 0.6 내외의 값이 나왔다. 이는 본 연구에서 명료성과 참신성의 일관성 계수가 각각 0.367, 0.258로 매우 낮게 나온 것과 상반된다. 주목할 만한 것은 배수정과 박주용의 연구에서는 학생들에게 주별로 주어진 자료를 읽은 후 그와 관련된 ‘글쓰기’ 과제를 해결하도록 하였다는 점이다. 본 연구에서는 학생들에게 ‘증명 작성’을 하도록 하였다는 점과 차이가 있다. 앞서 II절에서 다루었듯이 증명 작성은 일반적인 글쓰기와 구별되는 특수성을 지닌다. 대학생들은 학창시절부터 일반적인 글을 읽고 쓰는 경험을 지속적으로 해왔으므로, 글쓰기 능력을 기를 기회가 많이 있었고 따라서 한 학기 동안 학생들의 글쓰기 능력이 크게 변화했을 것이라고 보기는 어렵다. 그러나 본 연구의 참여자들은 대부분 학부 저학년 학생들로 엄밀한 수학적 증명 작성을 배우기 시작하는 과정에 있는 학생들이었다. 또한 증명 작성을 배우는 것이 학부 수학 전공 수업들의 주요 목표이기도 하다. 따

라서 연구에 참여한 학생들의 증명 작성 능력이 일관적이었을 것이라고 보기에는 무리가 있으며, 이러한 점이 낮은 일관성 계수가 도출된 원인으로 보인다.

2. 타당도

동료평가의 타당도를 조사하는 방법 중 하나로 학생들이 얼마나 전문가와 유사한 평가를 하는지를 조사하는 방법이 있다. 본 연구에서는 증명 동료평가의 타당도를 조사하기 위해 〈정수론〉 수강생들과 전문가를 대상으로 '증명 평가과제 1'과 '증명 평가과제 2'를 실시하였다.[11] 이번 절에서는 수강생 집단과 전문가 집단 간의 증명 평가과제 점수의 상관을 조사함으로써 증명 동료평가의 타당도를 살펴본다. 타당도 검사를 위해 측정한 증명 평가과제의 학생-전문가 평가 점수의 상관 계수는 다음 〈표 Ⅳ-16〉과 같다.

〈표 Ⅳ-16〉 학생-전문가 평가 점수의 상관 계수

학생	전체 점수	논리성	명료성	참신성
증명 평가과제 1	0.852	0.752	0.170	-0.018
증명 평가과제 2	0.824	0.628	0.856	0.745

측정 결과, '증명 평가과제 1'에서 전체 점수와 논리성 점수는 학생들의 채점 결과와 전문가의 채점 결과가 높은 정적 상관(상관계수 0.851)

11) 전문가가 작성한 '증명 평가과제 1'과 '증명 평가과제 1'의 응답 내용은 모두 부록에 수록하였다.

을 보였으나, 명료성 점수와 참신성 점수는 0에 가까운 상관계수(명료성 0.170, 참신성 -0.018)가 측정되었다. 이는, ‘증명 평가과제 1’에서 학생들의 논리성 채점은 전문가의 채점과 비슷하나, 명료성과 참신성은 전문가의 채점과 유사성이 거의 없다고 볼 수 있다. 명료성 점수와 참신성 점수의 유사성이 거의 없다고 나타난 한편, 전체 점수의 유사도는 매우 높았다. 즉, 학생들과 전문가가 영역별로는 차이가 나는 채점을 하지만 이들이 부여한 영역별 점수를 합하였을 때에는 유사한 결과를 보였다. 이는 학생들이 명료성이 떨어지고 참신성은 높다고 부여한 증명을 전문가는 명료성은 높으나 참신성이 부족하다고 부여하는 등의 현상으로 인한 것으로 보인다.

‘증명 평가과제 2’에서는 전문가의 채점결과와 학생들의 채점결과가 전체 점수뿐만 아니라 논리성 점수, 명료성 점수 그리고 참신성 점수 모두에서 높은 정적 상관을 보였다. 이러한 결과를 통해 학생들이 학기 말에 이르러서는 전문가와 유사한 평가를 하게 되었다고 할 수 있다. 전체 점수와 논리성 점수의 상관 계수가 소폭 감소하였는데 이는 ‘증명 평가과제 2’에서 다룬 정리와 증명이 포함하는 수학적 내용이 ‘증명 평가과제 1’에서 다룬 정리와 증명이 포함하는 수학적 내용에 비해 난이도가 높은 것에서 야기된 것으로 보인다.

‘증명 평가과제 1’에서 학생 집단과 전문가 집단 간에 논리성 평가는 유사도가 높았으나 명료성 평가와 참신성 평가의 유사도는 매우 낮았다. Selden과 Selden(2015)의 분류에 따르면, 본 증명 동료평가에서 논리성의 평가는 증명 검증에 해당하고 명료성과 참신성의 평가는 증명 평가에 해당한다. 이러한 분류에 따르면 학생들이 학기 중반에 증명 검증은 잘 수행하지만 증명 평가는 잘 수행하지 못하고 있다고 해석할 수 있다. 이는 Tao(2007)가 증명 평가 능력은 수학자들이 오랜 경험을 통해 기르

게 되는 것이라고 주장한 것과 일치하는 결과라고 볼 수 있다. 한편 학기말에 이르러 명료성과 참신성의 타당도가 높아진 것을 통해서 학생들의 증명 평가 능력이 크게 향상되었음을 확인할 수 있다. 이를 토대로 한 학기의 증명 검증 및 증명 평가 활동이 학생들에게 유의미한 경험을 제공하였을 것으로 볼 수 있다.

2.1. 타당도가 낮은 표본의 내용 분석

이 절에서는 ‘증명 평가과제 1’에서 학생 집단의 평가와 전문가 집단의 평가에 어떠한 차이가 있는지 구체적으로 살펴보도록 한다. 특히 평가결과가 집단에 따라 큰 차이를 보인 증명 1의 명료성 평가 결과, 증명 3의 명료성 평가 결과, 증명 4의 명료성과 참신성 평가 결과, 증명 5의 논리성과 명료성 평가 결과를 위주로 살펴본다.

<표 Ⅳ-17>은 전문가 집단의 ‘증명 평가과제 1’의 증명 1의 명료성 평가 결과를 나타낸 것이다.

<표 Ⅳ-17> 증명 평가과제 1의 증명 1 평가 결과(전문가/명료성)

	명료성 평가 근거	점수
전문가 1	“1. ①, ②, 그리고 ③은 소수들을 문자로 표현한다면 보다 간결하고 명확히 표현될 수 있다고 생각합니다. 2. ①에서 귀류법을 언급했으므로 ⑤는 의미가 없습니다.”	1
전문가 2	“2번에서 3번 문장을 굳이 나눠서 쓸 필요가 없을 것 같습니다. ‘그러면 가장 큰 소수 k가 존재한다.’ 는 표현이 조금 더 매끄러운 것 같습니다.”	2

전문가 2인은 해당 증명의 명료성에 각각 1점과 2점을 부여하여 서로

다른 점수를 부여하였으며 서로 다른 평가 근거를 들고 있었다. 전문가 1은 기호를 사용하지 않아 명료성이 떨어진다고 보았으며 필요하지 않은 문장이 있다고 언급하였다. 한편 전문가 2는 ‘가장 큰 소수가 존재함’을 주장하는 2번 문장과 ‘가장 큰 소수를 k’라고 상정하는 3번 문장을 ‘가장 큰 소수 k가 존재함’을 주장하는 문장 하나로 바꾸는 것이 보다 매끄러울 것이라고 언급하였다. 전문가들은 해당 논증이 어떤 점에서 명료하지 않은지를 언급하였을 뿐만 아니라 보다 명료한 논증이 되기 위해서 어떻게 수정하는 것이 좋을지도 서술하고 있었다. 다음으로 학생들은 같은 논증의 명료성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-18>은 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-18> 증명 평가과제 1의 증명 1 평가 결과(학생/명료성)

명료성 평가 근거	점수	학생 수
군더더기가 없음/명료함/간단함/불필요한 내용이 없음	3	14
알아보기 쉬움/이해하기 쉬움/받아들이기 쉬움	3	3
(근거 적지 않음)	3	2
굳이 ⑤를 쓸 필요가 없음	2	1

학생 집단은 95%(19명)가 3점을, 5%(1명)가 2점을 부여하였다. 명료성에 3점을 부여한 학생들은 “군더더기가 없다”, “명료하다”, “간단하다”, “불필요한 내용이 없다”등을 이유로 들었다. 2점을 부여한 한 학생은 “굳이 ⑤를 적을 필요가 없다”는 것을 이유로 들었다. 전문가 2인과 다르게 명료성에 3점을 부여한 학생들은 구체적인 설명 없이 평가

근거를 간략하고 주관적으로 설명하였다. 전문가 1인과 동일하게 2점을 부여한 학생은 “굳이 5번 문장을 적을 필요가 없다”고 언급하며 구체적인 근거를 들고 있었다. 대부분의 학생들이 주관적이고 간략한 근거를 서술하는 현상이 발견되었으며, 특별히 전문가와 다른 점수를 부여하는 경우에 그러한 현상이 두드러졌다.

계속해서 타당도가 낮게 나온 또 다른 표본을 살펴본다. 다음 <표 Ⅳ-19>는 전문가 집단의 ‘증명 평가과제 1’의 증명 3의 명료성 평가 결과를 나타낸 것이다.

<표 Ⅳ-19> 증명 평가과제 1의 증명 3 평가 결과(전문가/명료성)

	명료성 평가 근거	점수
전문가 1	“1. 지나치게 명료성만을 추구한 나머지, 증명에 포함되어야 할 내용이 다수 빠져있습니다. 2. 정의되지 않은 기호들을 사용했습니다.”	2
전문가 2	“일단 $\pi(n)$이 어떤 함수인지에 그리고 n이 자연수라는 사실이 정확히 명시되어야합니다. 그리고 오른쪽 화살표를 쓰는 방법보다는 정확히 limit을 이용하여 적어주는 것이 나아보입니다.”	1

전문가 2인은 해당 증명의 명료성에 각각 2점과 1점을 부여하여 서로 다른 점수를 부여하였으며 서로 다른 평가 근거를 들고 있었다. 전문가 1은 해당 논증에서 “증명에 포함되어야 할 내용이 다수 빠져있으며 정의되지 않은 기호들을 사용했다”고 언급하였다. 한편 전문가 2는 ‘$\pi(n)$이 어떤 함수인지’와 ‘n이 자연수라는 것’이 명시되어야 하며 “극한을 표기할 때에 화살표를 사용하는 것보다 limit를 사용하는 것이

나올 것”이라고 언급하였다. 전문가들은 해당 논증이 어떤 점에서 명료하지 않은지를 언급하였을 뿐만 아니라 보다 명료한 논증이 되기 위해서 어떻게 수정하는 것이 좋을지도 서술하고 있었다. 다음으로 학생들은 같은 논증의 명료성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-20>은 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-20> 증명 평가과제 1의 증명 3 평가 결과(학생/명료성)

명료성 평가 근거	점수	학생 수
명료함	3	5
실수를 했지만 증명 자체는 명료함	2	3
비약이 많음	2	2
로그를 사용하여 명확하지 않은 증명이 됨	2	1
(근거 적지 않음)	2	1
증명을 이해할 수 없음	1	3
논리성이 너무 떨어짐	1	3
2의 근거가 불충분함	1	2

학생들의 25%(5명)는 3점을, 35%(7명)는 2점을, 40%(8명)는 1점을 부여하였다. 3점을 부여한 학생들은 그 이유를 짧게 “명료하다”라고 서술하였다. 2점을 부여한 학생들은 “비약이 많다”, “로그를 사용하여 명확하지 않은 증명이 되었다”, “실수를 했지만 증명 자체는 명료하다” 등을 이유로 들었고, 1점을 부여한 학생들은 “증명을 이해할 수 없

다”, “논리성이 너무 떨어진다”, “②의 근거가 불충분하다” 등을 이유로 들었다. 증명 1에서의 결과와 동일하게, 전문가와 차이가 있는 평가를 한 5명의 학생들은 주관적인 이유를 적는 경향성이 발견되었다. 2점과 1점을 준 학생들 중 “비약이 많다”, “논리성이 너무 떨어진다”, “②의 근거가 불충분하다”를 이유로 지적한 학생들이 있음에 주목하자. 이 학생들은 명료성을 평가할 때에 해당 증명이 논리적인지, 논리적으로 결함이 없는지를 반영하였다고 볼 수 있다. 이들의 평가 내용을 통해 평가하는 증명의 논리성이 명료성의 평가에 영향을 줄 때도 있음을 확인할 수 있다.

계속해서 타당성이 낮게 나온 평가들, 즉 전문가와 학생들의 평가가 다소 큰 차이를 보이는 표본들을 살펴보도록 한다. ‘증명 평가과제 1’의 증명들 중에서도 특히 네 번째 증명의 명료성과 참신성에 관한 전문가들의 채점과 학생들의 채점이 큰 차이를 보였다. <표 Ⅳ-21>은 전문가 집단의 ‘증명 평가과제 1’의 증명 4의 명료성 평가 결과를 나타낸 것이다.

<표 Ⅳ-21> 증명 평가과제 1의 증명 4 평가 결과(전문가/명료성)

	명료성 평가 근거	점수
전문가 1	“②와 ③을 분리할 이유가 없습니다.”	2
전문가 2	“이 증명에서는 굳이 귀류법을 사용할 필요가 없습니다. ‘4k+3 꼴의 소수는 전체 소수 집합의 부분집합이다. 정리 2.38에 의해 이 부분집합이 무한하므로 소수의 집합도 무한하다.’ 와 같은 논리가 더 좋아 보입니다.”	2

전문가 2인 모두 해당 증명의 명료성에 2점을 부여하였다. 한편 이들은 서로 다른 평가 근거를 들었다. 전문가 1은 2번 문장과 3번 문장을 분리할 이유가 없다고 언급하였다. 전문가 2는 집합의 개념을 사용하여 ‘무한집합을 부분집합으로 갖는 집합은 무한집합’이라는 정리를 적용하면 귀류법을 쓰지 않고 직접 증명이 가능하므로 “굳이 귀류법을 사용할 필요가 없다”라고 언급하였다. 이 표본에서도 전문가들은 논증이 어떤 점에서 명료하지 않은지를 언급하였을 뿐만 아니라 보다 명료한 논증이 되기 위해서 어떻게 수정하는 것이 좋을지도 서술하고 있었다. 다음으로 학생들은 같은 논증의 명료성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-22>는 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-22> 증명 평가과제 1의 증명 4 평가 결과(학생/명료성)

명료성 평가 근거	점수	학생 수
깔끔함/명료함	3	10
수학적으로 필요한 내용만 딱 들어가 있음/군더더기가 없이 깔끔함	3	3
앞의 정리를 이용해 깔끔함	3	1
간단함	3	1
③ 에서 사용한 정리를 정명하지 않았다면 사용할 수 없음	3	1
(근거 적지 않음)	3	2
명료하나 논리적으로 심각한 오류가 있음	2	1
③에서 귀류법을 쓰지 않아도 자명하게 무한함을 설명할 수 있음	2	1

전문가 평가결과와 대조적으로 90%에 해당하는 18명의 학생들은 명료성에 만점인 3점을 부여하였다. 전문가 2인과 다르게 3점을 부여한 학생들의 대부분이 “깔끔함”, “필요한 내용만 들어가 있음”, “간단함” 등 간략하고 주관적인 근거를 들었을 뿐 구체적인 설명은 하지 않았다. 반면 전문가 2인과 동일하게 2점을 매긴 학생 2명 중 1명은 “귀류법을 쓰지 않아도 무한함을 설명할 수 있다”라고 보다 구체적인 근거를 서술하였다. 그러나 전문가 2는 직접 증명을 위한 구체적인 아이디어를 설명하였던 것과 달리 어떠한 아이디어로 직접 증명을 할 수 있는지는 설명하지 않았다.

계속해서 ‘증명 평가과제 1’의 증명 4의 참신성 평가결과를 살펴본

다. 다음 <표 Ⅳ-23>은 전문가 집단의 ‘증명 평가과제 1’의 증명 4의 참신성 평가 결과를 나타낸 것이다.

<표 Ⅳ-23> 증명 평가과제 1의 증명 4 평가 결과(전문가/참신성)

	참신성 평가 근거	점수
전문가 1	“위 증명에서 이용한 정리는 소수의 무한성 및 특정 형태의 소수 역시 무한함을 밝히고 있어 보이고자 하는 대상보다 더 강력합니다.”	1
전문가 2	“‘소수의 개수가 무한하다’를 증명하기 위해 이보다 더 강력한 정리를 사용하였으므로 별로 참신하다고 볼 수 없습니다.”	1

평가 결과를 보면 전문가 2인 모두 해당 증명의 참신성에 1점을 부여하였으며 동일한 근거를 들고 있었다. 전문가 2인 모두 해당 논증에서 보이고자 하는 정리보다 강력한 정리를 사용하였으므로 참신하다고 보기 어렵다고 서술하였다. 이처럼 전문가들은 참신성 평가의 근거를 구체적으로 서술하고 있었다. 다음으로 학생들은 같은 논증의 참신성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-24>는 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-24> 증명 평가과제 1의 증명 4 평가 결과(학생/참신성)

참신성 평가 근거	점수	학생 수
4k+3꼴을 생각한 것이 참신함/Theorem 2.38을 쓴 것이 참신함	2	3
쉽게 나올 수 있는 풀이가 아님/남들이 잘 하지 않는 풀이임	2	2
새로운 접근 방법	2	1
침소봉대 논리임	2	1
(근거 적지 않음)	2	2
일반적임	1	4
기존정리만 이용함/그냥 정리 사용한 것임	1	3
기존정리 활용했고 그 증명까지 들어갔다면 참신할 것임	1	1
(근거 적지 않음)	1	3

전문가 2인 모두 해당 논증이 ‘참신하지 않다’고 평가하였으나 학생들 중 45%에 해당하는 9명이 해당 논증이 ‘참신하다’고 평가하였다. 그 학생들은 해당 논증이 “남들이 잘 하지 않는 풀이다”, “참신하다”, “정리 2.38을 이용한 것이 참신하다” 는 것을 근거로 들었다. 해당 논증이 ‘참신하지 않다’고 평가한 학생들 중 일부는 “기존 정리를 사용한 것에 불과하므로 참신하지 않다”고 서술하였다. 이들은 다소 구체적인 근거를 들고 있었으나 “보다 강력한 정리를 사용했으므로 참신하지 않다”고 서술한 전문가의 평가 근거와는 차이가 있었다.

마지막으로 ‘증명 평가과제 1’의 증명 5의 평가 결과를 살펴본다.

다음 <표 Ⅳ-25>는 전문가 집단의 ‘증명 평가과제 1’의 증명 5의 논리성 평가 결과를 나타낸 것이다.

<표 Ⅳ-25> 증명 평가과제 1의 증명 5 평가 결과(전문가/논리성)

	논리성 평가 근거	점수
전문가 1	“1. ①, ④의 표현 ‘소수가 유한하다 ‘와 ‘소수가 무한하다’ 는 잘못되었습니다. 2. ②에서 나타난 n이 정의되지 않았습니다. 3. ③에서 새로 정의된 수가 소수임을 주장하는 과정에서, 소수의 정의를 이용하여 보이지 않았습니다.”	1
전문가 2	“3번 문장에 보충해야할 부분이 필요합니다. n개의 소수가 그 수를 나누지 못할 때 그 수가 왜 새로운 소수가 되는지에 대해 설명해야합니다. 이 부분은 귀류법 가정이 모순인데 핵심 부분이라고 생각합니다.”	3

‘증명 평가과제 1’의 증명 5의 평가에서 전문가 2인은 논리성을 각각 1점과 3점을 부여하였으며 평가 근거는 차이가 있었으나 서로 일치하는 부분도 있었다. 전문가 1은 ‘소수가 유한하다’와 ‘소수가 무한하다’를 ‘소수의 개수가 유한하다’와 ‘소수의 개수가 무한하다’는 표현으로 바꾸어야 한다고 언급하였으며 n이라는 수가 정의되지 않았음을 지적하였고 3번 문장에서 논리적인 결함이 있음을 지적하였다. 전문가 2는 ‘3번 문장에서 부연 설명이 필요함’을 지적하였다. 전문가 1과 전문가 2는 서로 다른 논리성 점수를 부여하였으나 어떠한 점에서 논증이 논리적이지 못한지를 구체적으로 언급하고 있었다. 다음으로 학생들은

같은 논증의 논리성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-26>은 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-26> 증명 평가과제 1의 증명 5 평가 결과(학생/논리성)

논리성 평가 근거	점수	학생 수
논리적 하자 없음/논리적임	5	9
(근거 적지 않음)	5	2
③ 에서 그래서 소수가 됨을 이끌어 내는데 비약이 있음	3	5
③ 에서 산술의 기본 정리를 언급해야 함	3	1
③ 에서 n개의 소수가 그 수를 나누지 못함을 설명해야 함	3	1
③ 에서 소수가 n개가 전부라는 것에 모순인 것	3	1
(근거 적지 않음)	3	1

전문가들이 해당 논증의 논리성 영역에 각각 1점과 3점을 부여한 것에 반해 절반이 넘는 55%(11명)의 학생들이 5점을 부여하였다. 나머지 45%(9명)의 학생들은 3점을 부여하였다. 전문가와 차이를 보이며 논리성 영역에 5점을 부여한 학생들은 "논리적 하자가 없음", "논리적임" 과 같이 간략하고 주관적인 근거를 들고 있었다. 반면 전문가 1인과 동일하게 3점을 부여한 학생 9명 중 8명이 ③에서 논리적인 비약이 있음을 지적하며 보다 구체적인 근거를 들고 있었다. 이러한 지적은 전문가 2인이 적은 근거와 일치하는 부분이었다.

다음으로 ‘증명 평가과제 1’의 증명 5 명료성 평가 내용을 살펴본다. <표 Ⅳ-27>는 전문가 집단의 ‘증명 평가과제 1’의 증명 5의 명료성 평가 결과를 나타낸 것이다.

<표 Ⅳ-27> 증명 평가과제 1의 증명 5 평가 결과(전문가/명료성)

	명료성 평가 근거	점수
전문가 1	“문자를 사용하지 않은 이유로, ③에서 같은 대상을 지칭하는 표현이 다릅니다.”	2
전문가 2	“증명을 서술할 때 적절한 기호를 사용하지 않으면 혼란스러울 수 있습니다. 그 수라는 표현보다는 증명 2) 에서와 같은 표현이 더 명확합니다.”	2

‘증명 평가과제 1’의 증명 5의 평가에서 전문가 2인 모두 명료성에 2점을 부여하였으며 유사한 근거를 제시하였다. 전문가 1과 전문가 2 모두 해당 논증에서 문자를 사용하지 않아 읽는 이에게 다소 혼란을 줄 수 있음을 지적하였다. 또한 전문가 1은 3번 문장에서 같은 대상을 지칭하는 표현이 다르다는 것을 언급하였으며 전문가 2는 “그 수”라는 표현을 사용하는 것보다는 기호를 사용하는 것이 더 명확하다고 설명하였다. 전문가 2인은 해당 논증이 어떠한 점에서 명료하지 못한지를 구체적으로 설명하고 있었다. 다음으로 학생들은 같은 논증의 명료성을 어떻게 평가하고 있는지 살펴보도록 한다. 다음 <표 Ⅳ-28>은 동일한 증명에 대해 학생들이 평가한 결과를 나타낸 것이다.

<표 Ⅳ-28> 증명 평가과제 1의 증명 5 평가 결과(학생/명료성)

명료성 평가 근거	점수	학생 수
명료함/깔끔함	3	5
한글이 많지만 너저분하지는 않음/한글이 많으나 불필요한 표현은 없음	3	3
군더더기 없는 설명임/불필요한 파트가 없음	3	2
(근거 적지 않음)	3	1
유한개의 소수를 기호로 표현하여 서술되면 좋을 것 같음/수학적 기호 표현이 필요함/소수 명시 필요함	2	6
글로만 써져 있어 알아보기 힘듦	2	1
글로만 써져있기도 하고 ④를 쓸 필요는 없음	2	1
증명 2와 같은 이유	2	1

전문가 2인 모두 해당 논증의 명료성에 2점을 부여한 것에 반해 학생들의 55%(11명)는 3점을 부여하였다. 나머지 45%(9명)의 학생들은 2점을 부여하였다. 전문가와 차이를 보이며 3점을 부여한 학생들은 “군더더기 없는 설명이다”, “명료하다”, “깔끔하다”, “불필요한 파트가 없다” 등 주관적인 근거를 들었다. 전문가와 동일하게 2점을 부여한 학생들은 “유한개의 소수를 기호로 서술하면 좋을 것 같다”, “수학적 기호 표현이 필요하다”, “④를 쓸 필요가 없다” 등 보다 구체적인 근거를 들었다.

지금까지 ‘증명 평가과제 1’에서 전문가와 학생 집단 간의 평가 결과에 큰 차이를 보이는 평가들을 구체적으로 살펴본 결과, 전문가들은 구체적인 근거를 바탕으로 평가를 하는 반면 학생들이 주관적인 근거로 평

가를 하는 현상이 나타났다. 또한 전문가의 평가와 유사도가 떨어질수록 학생들이 주관적인 근거를 토대로 평가하는 현상이 더 크게 나타났다. 이와 더불어 논리성 평가의 근거를 토대로 명료성 평가를 하는 현상도 나타났다. 이러한 현상들이 학기 중반에 학생들의 평가와 전문가의 평가 간에 낮은 유사성을 야기한 것으로 보인다.

2.2. 타당도가 높은 표본의 내용 분석

이상에서는 타당도가 낮은 표본들의 평가 결과를 분석하여 타당도가 낮은 표본들에서 학생들의 평가가 어떠한 특징을 보이는지를 살펴보았다. 이 절에서는 학생 집단과 전문가 집단 간에 채점 유사도가 높았던 ‘증명 평가과제 2’의 증명들을 중심으로 학생들과 전문가들이 어떠한 근거로 증명의 논리성과 명료성, 그리고 참신성을 평가하는지 살펴보도록 한다.

다음 <표 Ⅳ-29>는 전문가 집단의 ‘증명 평가과제 2’의 증명 3의 논리성 평가 결과를 나타낸 것이다.

<표 Ⅳ-29> 증명 평가과제 2의 증명 3 평가 결과(전문가/논리성)

	논리성 평가 근거	점수
전문가 1	“논리적으로 문제가 없는 증명이라 생각합니다.”	5
전문가 2	“각 문장사이에 논리적인 비약이나 오류는 전혀 보이지 않습니다.”	5

‘증명 평가과제 2’의 증명 3의 평가에서 전문가 2인 모두 논리성에 5점을 부여하였다. 이들은 “논리적으로 문제가 없다”, “논리적인 비

약이나 오류는 보이지 않는다”는 것을 근거로 제시하였는데 이들의 근거는 간략했다. 2.1절에서 타당도가 낮은 표본들을 살펴보았을 때에는 전문가들이 구체적인 근거를 들고 있었던 것과 다르게 근거 서술이 구체적이지 않았다. 이는 논리적인 결함이 없는 논증의 논리성을 평가할 때에 나타나는 현상이라고 볼 수 있다. 논리적인 결함이 있는 증명의 경우에는 어떤 부분에 논리적인 결함이 있는지 설명하고 개선 방향을 제시할 수 있는 한편 논리적인 결함이 없는 증명의 경우에는 ‘논증이 논리적이다’ 또는 ‘논리적인 결함이 없다’ 등의 언급 외에는 평가 근거를 작성하기가 어렵기 때문이다. 다음으로 같은 논증의 논리성을 학생들은 어떻게 평가하였는지 살펴본다. 다음 <표 Ⅳ-30>은 학생 집단이 동일한 증명의 논리성을 평가한 결과다.

<표 Ⅳ-30> 증명 평가과제 2의 증명 3 평가 결과(학생/논리성)

논리성 평가 근거	점수	학생 수
논리적인 결함이 없음/논리적임	5	4
정리 4.16을 증명한 상태라면 논리적인 문제가 없음	5	4
완벽한 것 같음	5	1
“수업 자료에 보면 페르마 소정리(ver. 1)과 페르마 소정리(ver. 2)가 서로 동치라고 나와있었던 것 같습니다. 위의 증명 3)이 제대로 된 논리적인 증명이 되기 위해서는 정리 4.16 자체를 아예 증명하는 방향으로 이끌고 가는 과정이 ①에서 ②번 과정 사이에 존재했어야 한다고 생각합니다.”	3	1
3번 문장에서 나누면 이라는 용어가 적절치 않음	3	1
순환논리가 사용되었음	1	2
“정리 4.16또한 페르마의 소정리입니다. 위의 증명은 페르마의 소정리 두 가지 버전이 동치라는 것의 한쪽 방향을 증명한 것입니다.”	1	1

참여 학생 14명 중 64.3%에 해당하는 9명이 전문가와 동일하게 논리성에 5점을 부여하였다. 이 9명의 학생들도 전문가와 마찬가지로 논리적인 결함이 없음을 근거로 들었다. 해당 평가에서는 전문가들과 대부분의 학생들이 구체적인 근거를 들지 않았는데 이는 해당 논증에 논리적인 결함이 없었고 때문에 전문가 2인과 대부분의 학생들이 논리성에 만점을 부여했다는 점에서 기인한 것으로 보인다. 논리성에 만점을 준 경우에는 특별히 지적하거나 근거로 들 만한 것이 없을 것이기 때문이다.

다음으로 ‘증명 평가과제 2’의 증명 3의 참신성 평가 결과를 살펴본

다. <표 Ⅳ-31>은 전문가 집단의 ‘증명 평가과제 2’의 증명 3의 참신성 평가 결과를 나타낸 것이다.

<표 Ⅳ-31> 증명 평가과제 2의 증명 3 평가 결과(전문가/참신성)

	참신성 평가 근거	점수
전문가 1	“정리 4.16이 보다 일반적인 내용을 다루고 있다고 생각합니다.”	1
전문가 2	“이 증명은 기초정수론에서 전형적으로 사용하는 방법을 나열했을 뿐입니다. 특별히 뛰어나다고 할 만한 아이디어가 사용되지 않은 것 같습니다.”	1

전문가들은 해당 논증의 참신성에 1점을 부여하였으며 본 논증이 참신하지 않다고 평가한 근거를 구체적으로 서술하고 있었다. 전문가 1은 보다 일반적인 내용을 다루는 4.16을 사용하였으므로 참신하다고 볼 수 없다고 서술하였다. 전문가 2는 해당 논증이 기초정수론에서 사용하는 전형적인 방법들을 사용하였기 때문에 참신하다고 볼 수 없다고 설명하였다. 전문가 2인 모두 해당 논증이 참신하지 않다고 생각하는 이유를 구체적으로 제시하고 있었다. 계속해서 동일한 증명의 참신성에 대해 학생들이 평가한 것을 살펴본다. 다음 <표 Ⅳ-32>는 학생 집단이 동일한 증명의 참신성을 평가한 결과다.

<표 Ⅳ-32> 증명 평가과제 2의 증명 3 평가 결과(학생/참신성)

참신성 평가 근거	점수	학생 수
일반적임/보편적임/평범함	1	5
참신하지 않다	1	1
새롭지 않다	1	1
누구나 생각할 수 있음/가장 쉽게 생각할 수 있는 풀이임	1	4
올바르지 못한 증명이기에 참신하다고 할 수 없음	1	1
직관적인 면이 떨어짐	1	1
위의 증명들과 동일한 이유	1	1

학생들 모두가 전문가 2인과 동일하게 참신성에 1점을 부여하였다. 학생들 모두 “일반적이다”, “참신하지 않다” 등 간략하고 주관적인 근거를 제시하고 있었다. 전문가가 구체적인 근거를 제시한 것과 차이를 보였다.

다음으로 ‘증명 평가과제 2’의 증명 4의 참신성 평가 결과를 살펴보도록 한다. 다음 <표 Ⅳ-33>은 전문가 집단의 ‘증명 평가과제 2’의 증명 4의 참신성 평가 결과를 나타낸 것이다.

<표 Ⅳ-33> 증명 평가과제 2의 증명 4 평가 결과(전문가/참신성)

	참신성 평가 근거	점수
전문가 1	“이항정리를 활용한 귀납법에 의한 풀이가 독창적이라고 생각합니다.”	2
전문가 2	“기존에 잘 알려진 페르마 정리의 증명과는 다르게 귀납법을 적용하였습니다. 새로운 아이디어를 제시했다는 점에서 참신하다고 생각합니다.”	2

전문가 2인 모두 해당 논증의 참신성에 2점을 부여하였다. 전문가 1은 “이항정리를 활용한 귀납법에 의한 풀이가 독창적”이라는 근거를 제시하였고 전문가 2는 “기존에 잘 알려진 증명과 다르게 귀납법을 적용한 것이 참신하다”고 평가하였다. 전문가 2인 모두 해당 논증이 참신하다고 평가한 근거를 구체적으로 서술하고 있었다. 계속해서 동일한 논증의 참신성에 대해 학생들이 작성한 평가를 살펴보도록 한다. 다음 <표 Ⅳ-34>은 학생 집단이 동일한 증명의 참신성을 평가한 결과다.

<표 Ⅳ-34> 증명 평가과제 2의 증명 4 평가 결과(학생/참신성)

참신성 평가 근거	점수	학생 수
새로운 방법임/처음 보는 방법	2	5
떠올리기 어려움/생각해본 적 없는 풀이임	2	2
페르마 소정리2를 증명함으로써 접근하고자 한 것이 참신함	2	1
논리적으로 틀렸으므로 참신하다고 할 수 없음	1	3
누구나 생각할 수 있는 방법	1	2
일반적임	1	1

14명의 참여 학생 중 57.1%에 해당하는 8명의 학생들이 참신성에 2점을 부여하였고 나머지 5명은 1점을 부여하였다. 참신성에 2점을 부여한 8명 중 7명의 학생들이 “새로운 방법이다”, “떠올리기 어렵다” 등 주관적이고 간략한 근거를 제시하였다. 참신성에 2점을 부여한 8명 중 나머지 1인만이 전문가들과 같이 구체적인 근거를 들었다.

다음으로 ‘증명 평가과제 2’의 증명 5의 논리성 평가 결과를 살펴보도록 한다. 다음 <표 Ⅳ-35>은 전문가 집단의 ‘증명 평가과제 2’의 증명 5의 논리성 평가 결과를 나타낸 것이다.

<표 Ⅳ-35> 증명 평가과제 2의 증명 5 평가 결과(전문가/논리성)

	논리성 평가 근거	점수
전문가 1	“②는 소수의 정의가 아닌 성질입니다.”	5
전문가 2	“각 문장 사이에 논리적인 비약이나 오류가 보이지 않습니다.”	5

전문가 2인 모두 해당 논증의 논리성에 만점인 5점을 부여하였다. 전문가 1은 비록 논리성에 만점을 주었으나 2번 문장에 오류가 있음을 지적하였다. 전문가 2는 “각 문장 사이에 논리적인 비약이나 오류가 보이지 않음”을 근거로 제시하였다. 앞서 살펴본 ‘증명 평가과제 2’의 증명 3의 논리성 평가 결과와 마찬가지로 논리성에 만점을 부여하는 경우라서 구체적인 평가 근거를 작성하지 않는 현상이 나타났다고 볼 수 있다. 계속해서 학생들이 작성한 평가 결과를 살펴본다. 다음 <표 Ⅳ-36>은 학생 집단이 동일한 증명의 논리성을 평가한 결과다.

<표 Ⅳ-36> 증명 평가과제 2의 증명 5 평가 결과(학생/논리성)

논리성 평가 근거	점수	학생 수
논리적인 결함이 없음/문제가 없음/비약이 없음	5	3
논리적임	5	3
오일러 정리를 먼저 얻을 수 있다면 논리적인 문제가 없음	5	3
오일러의 정리를 이용하여 잘 증명하였음	5	1
완벽함	5	1
"먼저, Suppose that p is a prime and a is an integer relatively prime to p. 라는 문장이 들어가야 하는데 들어가지 않았습니다. (-1점) ②에서, 사실 수업자료에서는 소수의 정의를 ②에서와 같이 다루지 않았기 때문에 정의에 의해 ②와같은결론이나왔다고하는점에서잠시보류해야하는부분이 있다고 생각합니다.(-0.2점) 이렇게 평가하면 3.8점이 나오게 되는데 올림평가(?)하게 되면 5점을 줄 수 있게 됩니다."	5	1
"2번에서 '어떤'이 아니라 '임의'의로 수정해야한다."	3	1
"2번에서 '어떤'이 아니라 '모든'이 되어야 할 것이다. 또한 이번 정수론 강의 교육과정상 오일러정리는 아직 증명되지 않았다."	1	1

14명의 참여 학생들 중 78.6%인 11명의 학생들이 전문가와 동일하게 논리성에 5점을 부여하였다. 만점을 부여한 전문가들과 학생들 대부분이 해당 논증이 “논리적임”, “논리적인 결함이 없음” 등 주관적이고 추상적인 이유를 근거로 제시하였다. 논리성에 3점 또는 1점을 준 학생들은 2번에서 ‘어떤’이라는 표현을 ‘모든’으로 고쳐야 한다고 지적하는 등 구체적인 근거를 제시하였다. ‘증명 평가과제 2’의 증명 3의 논리성 평가 결과와 마찬가지로 평가자들이 논리성에 만점을 부여하는 경우에 평가 근거를 간략하게 작성하는 현상이 나타났다.

계속해서 타당성이 높게 나왔던 ‘증명 평가과제 2’의 증명 5의 참신성 평가 결과를 살펴본다. <표 Ⅳ-37>는 전문가 집단의 ‘증명 평가과제 2’의 증명 5의 참신성 평가 결과를 나타낸 것이다.

<표 Ⅳ-37> 증명 평가과제 2의 증명 5 평가 결과(전문가/참신성)

	참신성 평가 근거	점수
전문가 1	“Euler’s Theorem은 보이고자 하는 정리의 일반화입니다.”	1
전문가 2	“위 방법은 페르마 정리보다 강력한 정리에서 특수한 경우만 관찰을 한 것입니다. 새로운 아이디어를 제공하지 않았습니다.”	1

전문가 2인 모두 해당 논증의 참신성 평가에서 1점을 부여하였으며 두 사람 모두 보다 강력한 정리의 특수한 경우를 관찰한 것에 불과하므로 해당 논증이 참신하다고 볼 수 없음을 근거로 제시하였다. 전문가들은 논증의 참신성을 평가함에 있어서 해당 논증이 정수론에서 어떤 의미를 가지는지를 고민하였고 이를 통해 전문가들이 의미론적(semantic) 의미에

집중하였다고 할 수 있다. 계속해서 학생들이 동일한 논증의 참신성에 대해 평가한 결과를 살펴본다. 다음 <표 Ⅳ-38>은 학생 집단이 동일한 증명의 참신성을 평가한 결과다.

<표 Ⅳ-38> 증명 평가과제 2의 증명 5 평가 결과(학생/참신성)

참신성 평가 근거	점수	학생 수
페르마 소정리는 오일러정리의 특수한 경우임을 보여주므로 참신함	2	1
일반적인 증명/보편적인 풀이/무난함	1	5
누구나 생각할 수 있는 풀이	1	3
참신하지 않음	1	2
당연한 결과	1	1
다른 정리를 이용하여 증명함	1	1
(근거 적지 않음)	1	1

학생들의 평가 결과를 보면 14명의 학생들 중 13명이 전문가 2인과 동일하게 참신성에 1점을 부여하였다. 나머지 1인은 2점을 부여하였다. 참신성에 1점을 부여한 학생들은 대부분 “일반적인 증명이다”. “참신하지 않다”와 같이 주관적이고 간략한 근거를 제시하였다. 참신성에 2점을 부여한 학생의 경우에는 본 논증이 페르마 소정리가 오일러 정리의 특수한 경우임을 보여주는 것이 참신하다고 평가하였다.

타당도가 높은 표본들을 살펴본 결과 타당도가 낮은 표본들과 마찬가지로 전문가들은 구체적인 근거를 바탕으로 평가를 하는 반면 학생들은 주관적이고 추상적인 근거를 바탕으로 평가를 하는 현상이 나타났다. 해

당 표본들에서 학생들이 준 점수와 전문가들이 준 점수가 높은 일치도를 보이지만 전문가들과 학생들이 근거로 삼은 것에는 큰 차이가 있었던 것이다. 비록 학기말에 이르러서 학생들이 전문가와 유사한 점수를 부여하게 된 한편 학생들이 증명을 평가할 때에 제공하는 피드백이 여전히 추상적임을 의미한다. 이러한 점을 통해 증명 동료평가를 통해 피채점자에게 주어지는 피드백 내용이 전문가 평가를 통해 학생들에게 주어지는 피드백 내용에 비해 추상적일 것이라고 예상할 수 있다.

V. 결론

1. 요약

대학 수학교육에서 증명의 학습은 중요한 학습 목표로 여겨져 왔다. 한편 중고등학생 뿐 아니라 대학생들도 증명 학습에 있어서 많은 어려움을 겪는다는 보고가 있어왔다. 이러한 배경 속에서 학생들의 증명 학습을 돕기 위한 여러 가지 노력이 있어왔다. 특별히 주목할 점은 학생들을 수학적 의사소통에 참여하게 하였을 때에 학생들이 의미 있는 증명 학습을 한다는 연구 결과이다. 수업 시간에 학생들로 하여금 수학적 의사소통에 참여하도록 하는 방법으로 발표와 토론 활동이 있다면 교실 밖에서 학생들이 참여할 수 있는 수학적 의사소통 활동은 동료평가 활동이 있다. 실제로 동료평가는 글쓰기 과제에 적용하였을 때에 다양한 효과를 가진다는 것이 확인되었다.

이러한 맥락에서 증명 학습을 위한 새로운 교수학습 도구로써 증명 동료평가의 도입을 고려해볼 수 있다. 동료평가는 오랜 시간 활용되어온 도구임에도 불구하고 증명 과제에 동료평가를 적용한 사례는 찾아보기 어렵다. 그 이유는 무엇일까? 학생들의 배경지식이나 평가 경험이 부족 때문에 교수자와 학생들 모두 동료평가 결과를 신뢰할만한지 확신하지 못하는 것이 가장 큰 이유일 것이다. 따라서 증명 동료평가의 적극적인 도입을 주장하기에 앞서 증명 동료평가의 신뢰도와 타당도를 살펴보는 것이 필요하다. 이 연구의 목적은 온라인에서 이루어진 증명 동료평가의 신뢰도와 타당도를 조사함으로써 증명 동료평가의 활용 가능성을 살펴보는 것이다. 이러한 목적을 이루기 위한 연구질문은 다음과 같다.

1. 대학 정수론 수업에서 실시한 온라인 증명 동료평가의 신뢰도는 어떠한가?

2. 대학 정수론 수업에서 실시한 온라인 증명 동료평가의 타당도는 어떠한가?

위 두 가지 연구 질문에 대답하기 위하여 대학 <정수론> 수업에서 주별로 온라인 동료평가를 실시하고 이에 대한 연구를 진행하였다. 연구 질문 1번에 대답하기 위해서는 증명 동료평가의 신뢰도를 일치도와 일관성 면에서 살펴보았다. 일치도는 같은 답안을 서로 다른 평가자가 평가하였을 때에 얼마나 비슷한 결과가 나오는 지를 의미한다. 평가자들 간의 평가결과가 유사할수록 일치도가 높은 평가라고 할 수 있다. 일치도 조사를 위해서 학생들이 주별로 수행한 동료평가 결과 및 추가로 수집한 증명 평가과제의 결과를 토대로 학생 채점자간 채점 일치도를 살펴보았다. 일관성은 같은 피평가자의 능력에 대한 서로 다른 평가가 얼마나 일관성 있는가를 의미한다. 서로 다른 시점에 이루어진 평가들의 평가 결과가 유사할수록 일관성이 높은 평가라고 할 수 있다. 일관성 조사를 위해서 각 학생들이 한 학기 동안 받은 점수의 일관성을 조사하였다. 한편, 본 연구에서는 동료평가의 타당성을 조사하는 방법으로 동료평가 결과와 전문가 평가 결과가 얼마나 유사한지를 살펴보는 방법을 채택하였다. 따라서 연구 질문 2번에 대답하기 위해서 학생들이 수행한 증명 동려평가 결과와 전문가들이 같은 증명에 대해 평가한 결과가 얼마나 유사한지를 조사하였다.

먼저 일치도 측면에서 신뢰도 검사를 위해, 학생들이 주별로 수행한 증명 동료평가 결과를 토대로 채점자간 일치도를 검사하였다. 총 143회의 일치도 검사 결과, 0.6 미만의 값이 3회(0.214, 0.316, 0.500), 0.6 이상

0.7 미만의 값이 2회(0.667, 0.667), 0.7 이상 0.8 미만이 2회(0.750, 0.789) 0.8 이상 0.9 미만이 7회(0.846, 0.851, 0.861, 0.885, 0.889, 0.892, 0.892) 나온 것을 제외한 129회의 검사에서 0.9 이상의 값이 도출되었다. 이러한 결과는 하나의 증명에 대한 세 명의 평가자들이 준 점수들 간이 매우 유사함을 의미한다. 신뢰도 검사에서 학기말에 일치도가 아주 낮았던 데이터들(0.214, 0.316, 0.500)이 있었는데 이 데이터들은 각각 12주차, 11주차, 10주차에 이루어진 동료평가 결과 자료이다. 학기말에 이르러 일치도가 낮은 표본이 나타나는 현상은 학기 후반에 이르면서 다루는 수학적 내용이 어렵고 복잡해졌고, 이에 따라 한 정리에 대한 다양한 증명이 가능해졌으며 평가자가 자신이 평가하는 증명의 내용을 제대로 이해하지 못하는 경우가 생겼기 때문으로 보인다. 또한 채점 내용을 분석한 결과, 평가 과정에서 주관적인 판단이 수반되거나 평가자들 간에 배경지식이나 검증 능력의 차이를 나타낼수록 채점자간 일치도가 낮게 나오는 경향이 확인되었다. 채점하는 증명에 논리적인 결함이 발견되지 않거나, 특별히 명료하지 않다고 여길 부분이 없는 경우에 채점자는 논리성이나 명료성 점수를 몇 점을 부여해야할지 고민하지 않고 만점을 부여하게 된다. 비슷하게, 채점하는 증명이 일반적인 아이디어를 담고 있는 경우에 채점자는 참신성에 몇 점을 부여할지 고민하지 않아도 된다. 이러한 이유로 논리적인 결함이 없고, 명료하며 일반적인 증명의 경우에는 채점자의 주관적 판단이 필요하지 않아 채점자간 일치도가 매우 높게 도출되었다. 반대로 채점자의 주관적인 판단이 요구되는 경우 채점자간 일치도가 낮아졌다.

한편, 위 일치도 조사에서 사용된 표본들의 크기가 3으로 매우 작았다는 점에서 결과를 일반화하기 어려울 것으로 판단하고, 이러한 한계를 극복하기 위해 증명 평가과제를 통해 수집한 크기 20과 17의 표본들을

가지고 학생 집단의 채점자간 일치도를 다시 살펴보았다. 이 표본들을 가지고 채점자간 일치도를 조사한 결과, 학기 중반에 실시한 '증명 평가과제 1'과 학기말에 실시한 '증명 평가과제 2'에서 전체 점수 뿐 아니라 논리성, 명료성 그리고 참신성 점수 모두 매우 높은 채점자간 일치도를 보였다(증명 평가과제 1: 전체점수 0.943, 논리성 0.963, 명료성 0.943, 참신성 0.915, 증명 평가과제 2: 전체점수 0.970, 논리성 0.727, 명료성 0.854, 참신성 0.915). 이러한 결과를 통해 학생들이 학기 중반부터 학기말까지 서로 유사도가 높은 채점을 하고 있었음을 확인할 수 있다. 즉, 학기 중반부터 일치도 측면에서는 높은 신뢰도를 보였다.

다음으로 연구자는 학생들이 주별로 받은 동료평가 점수의 일관성을 조사하기 위하여 16명의 학생들이 매주 받은 네 개의 동료평가 점수의 평균을 구하였다. 일관성 조사에서는 점수들의 분산이나 상관계수 등이 주로 사용되는데 본 연구에서는 이때에 크론바흐 알파 테스트 사용해 상관계수를 측정하였다. 학생들이 한 학기 동안 받은 증명 동료평가 점수의 내적 일관성을 조사한 결과 그 일관성은 전체적으로 그리 높지 않았으며 특히 명료성 점수와 참신성 점수의 일관성은 낮은 편이었다(전체점수 0.620, 논리성 0.574, 명료성 0.367, 참신성 0.258). 평가에서 일관성을 조사할 때에는, 평가가 이루어지는 기간 동안 피평가자의 능력이 다소 일정하다는 가정이 수반된다. 그러나 본 연구에서 살펴본 동료평가가 진행되는 동안, 학생들은 증명에 대해 학습하는 시간을 가졌다. 따라서 각 학생들의 증명 작성 능력이 한 학기 동안 일정하다고 가정하는 것에는 무리가 있으며, 이러한 사실이 낮은 일관성을 야기했다고 볼 수 있다. 따라서 해당 표본에서는 낮은 일관성이 증명 동료평가의 낮은 신뢰도를 의미한다고 보기는 어렵다.

아무리 결과를 신뢰할 수 있는 평가도구라고 할지라도 타당성이 결여

된다면 해당 평가도구의 실효성을 주장하기는 어렵다. 특별히 동료평가에서는 아무리 학생들이 서로 일치도가 높은 평가를 하고 있을지라도 학생들의 평가결과가 전문가 평가 결과와 큰 차이를 보인다면 동료평가가 전문가 평가를 대체할 수 없을 뿐 아니라 동료평가의 결과가 타당하다고 볼 수 없다. 따라서 동료평가의 타당도 검사가 반드시 필요하다. 동료평가의 타당도를 조사하는 방법 중 하나로 학생들이 얼마나 전문가와 유사한 평가를 하는지를 조사하는 방법이 있다. 본 연구에서는 증명 동료평가의 타당도를 조사하기 위해 <정수론> 수강생들과 전문가를 대상으로 ‘증명 평가과제 1’과 ‘증명 평가과제 2’를 실시하였다. 학생들을 대상으로 한 ‘증명 평가과제 1’은 학기 중반에 이루어졌으며 ‘증명 평가과제 2’는 학기말에 이루어졌다. ‘증명 평가과제 1’에서 전체 점수와 논리성 점수는 학생들의 채점 결과와 전문가의 채점 결과가 높은 정적 상관을 보였으나, 명료성 점수와 참신성 점수는 0에 가까운 상관계수가 측정되었다. 이는, ‘증명 평가과제 1’에서 학생들의 논리성 채점은 전문가의 채점과 비슷하나, 명료성과 참신성은 전문가의 채점과 유사성이 거의 없다고 볼 수 있다. ‘증명 평가과제 2’에서는 전문가의 채점결과와 학생들의 채점결과가 전체 점수뿐만 아니라 논리성 점수, 명료성 점수 그리고 참신성 점수 모두에서 높은 정적 상관을 보였다. 다시 말해 학기 중반에는 전문가와 유사도가 낮은 증명을 하던 학생들이 학기 말에 이르러서는 전문가와 유사한 평가를 하게 된 것이다. 이러한 점을 통해 학생들이 학기 중에 전문가와 유사한 증명 검증 및 증명 평가 능력을 기르게 되었다고 할 수 있다.

왜 어떤 표본은 타당도가 높은 반면 다른 표본은 타당도가 낮은 것일까? 이 질문에 대답하기 위해 타당도가 낮게 나온 표본들과 타당도가 높게 나온 표본들의 내용을 분석하였다. ‘증명 평가과제 1’에서 타당도

가 낮게 나온 표본, 즉 전문가와 학생 집단 간의 평가 결과에 큰 차이를 보이는 표본들을 구체적으로 살펴본 결과, 전문가들은 구체적인 근거를 바탕으로 평가를 하는 반면 학생들이 주관적인 근거로 평가를 하는 현상이 나타났다. 또한 전문가의 평가와 유사도가 떨어질수록 학생들이 주관적인 근거를 토대로 평가하는 현상이 더 크게 나타났다. 이와 더불어 논리성 평가의 근거를 토대로 명료성 평가를 하는 현상도 나타났다. 계속해서 '증명 평가과제 2'에서 타당도가 높은 표본들, 즉 전문가 평가 결과와 학생들의 평가 결과가 유사한 표본들을 살펴본 결과, 타당도가 낮은 표본들과 마찬가지로 전문가들은 구체적인 근거를 바탕으로 평가를 하는 반면 학생들은 주관적이고 추상적인 근거를 바탕으로 평가를 하는 현상이 나타났다. 해당 표본들에서 학생들이 준 점수와 전문가들이 준 점수는 높은 일치도를 보인 한편 전문가들과 학생들이 근거로 삼은 것에는 큰 차이가 있었다. 타당도가 높아진 학기말에도 평가자 학생들이 제공하는 피드백이 여전히 추상적이었던 것이다. 학기말에도 여전히 평가를 받는 학생에게 큰 도움이 되지 않는 피드백이 제공되고 있었다고 할 수 있다. 이러한 현상은 학생들이 배경 지식과 평가 경험이 부족한 것에서 기인한 것으로 보인다.

2. 논의 및 제언

한 학기에 걸쳐 시행된 증명 동료평가 결과를 분석한 결과 일치도 측면의 신뢰도는 학기 전반에 걸쳐 매우 높았고 일관성 측면의 신뢰도는 그다지 높지 않았으며 타당도는 학기 중반에는 낮았으나 학기 말에 이르러서 높아졌다. 일치도의 경우에는 다루는 수학적 내용이 복잡해지고 어려워지는 학기말에 이르러서 약간 낮아지는 현상이 나타났다. 한편 학기 중반에 타당성이 낮게 나온 것은 학생들이 평가 경험이 부족하여 좋은

증명에 대한 견해가 전문가와 다소 차이를 보이기 때문으로 보인다. 이하에서는 본 연구의 결과가 가지는 함의를 살펴보고 여전히 해결되지 않은 증명 동료평가의 쟁점들을 다루도록 한다.

한 학기에 걸친 증명 동료평가에서 학생들은 전반적으로 신뢰도가 아주 높은 평가를 하고 있었다. 또한 학생들은 학기 초에는 논리성 평가에 있어서는 전문가와 유사한 수행을 보였으나 명료성과 참신성의 평가에 있어서는 전문가와 유사도가 낮은 수행을 하였다. 학생들은 학기 말에 이르러서는 논리성 영역 뿐 아니라 명료성과 참신성 영역에서도 전문가와 유사한 평가를 하게 되었다. Selden과 Selden(2015)의 분류에 따르면 학생들은 학기 중반에 전문가와 유사한 '증명 검증'을 하였으나 '증명 평가'에 있어서는 전문가와 다소 차이가 있는 수행을 하였고, 학기 말에 이르러서 '증명 검증' 뿐 아니라 '증명 평가'에서도 전문가와 유사한 수행을 하게 되었다. 학생들이 학기 말에 이르러 전문가와 비슷한 평가를 하게 된 것은 전문가와 유사한 증명 독해, 증명 검증, 증명 평가 역량을 가지게 되었음을 함의한다. 이처럼 학생들이 증명 독해, 증명 검증, 증명 평가 역량을 기르게 된 이유 또는 배경은 무엇일까? 이 질문에 대한 대답은 크게 두 가지가 가능하다. 첫 번째로는 학생들이 증명 동료평가에 참여함으로써 증명 독해, 증명 검증, 증명 평가에 대한 경험을 지속적으로 하게 되었고 이러한 경험이 학생들로 하여금 역량을 기르게 한 것으로 볼 수 있다. 가능한 또 다른 설명은 한 학기 동안 학생들이 수업을 들으며 증명에 대한 학습을 했다는 것이다. 학생들은 정수론 과목을 수강하면서 증명 독해와 증명 작성, 증명 검증 등의 활동에 참여하였다. 즉, 반드시 증명 동료평가가 아니더라도 학생들은 증명 학습의 기회가 있었기에 반드시 동료평가의 효과라고 보기 어렵다. 하지만 분명한 것은 다른 활동보다 동료평가 활동이 학생들로 하여금 적극적으

로 증명 검증과 증명 평가에 참여하도록 이끌었다는 것이다.

한편, 동료평가가 평가도구로써의 활용도를 갖기 위해서는 타당도의 확보가 필수적이다. 그러나 타당도 분석한 결과, 학기 초에 실시한 타당도 검사에서 학생 집단과 전문가 집단 간의 평가 결과의 유사도가 낮았다. 즉 학생들은 학기 초에 타당도가 낮은 평가를 하고 있었다. 이러한 점에서 증명 동료평가의 결과를 학점 산출을 위한 점수 계산에 포함시키는 것이 타당하다고 보기는 어렵다. 비록 학생들이 학기말에는 전문가와 유사한 즉, 타당도가 높은 평가를 하게 되었지만 학기 초의 평가 타당도가 낮았음을 부정할 수는 없다. 따라서 증명 동료평가의 결과를 실제 평가 점수로 활용할지, 학기 후반에 이루어진 점수만을 평가에 반영할지 등은 여전히 논란의 여지가 있다.

‘온라인’ 동료평가의 유익 중 하나는 평가자와 피평가자가 서로 알 수 없기 때문에 보다 높은 신뢰도를 확보할 수 있으며, 학생들이 겪는 심리적인 갈등의 많은 부분을 해결해준다는 것이다. 한편 수학적 증명은 보통의 글쓰기와 달리 수학적 기호를 포함하는 경우가 대부분이기 때문에 학생들이 수학적 기호나 수식을 입력하는 데에 어려움을 겪을 것으로 판단하고, 본 연구에서는 학생들에게 증명을 종이에 손으로 작성한 뒤 사진을 찍어 업로드하는 것을 허락하였다. 그런데 IBL 수업의 특성상 수업 시간에 학생 발표가 많이 이루어졌고, 이로 인해 학기말에 학생들이 서로의 글씨체를 알아보게 되면서 평가하는 증명이 누구의 증명인지 어느 정도 알게 되었던 것 같다. 실제로 6.17의 피드백 내용 중에 평가자 학생이 피평가자가 누구인지를 추측하고 실명을 언급한 경우도 있었다. 이처럼 손으로 작성한 증명을 사진으로 업로드하는 경우에 동료평가의 신뢰도가 다소 떨어지는 등의 문제점이 발생할 수 있다. 이러한 문제점은 온라인 동료평가 시스템에서 수식의 입력이 자유로워진다면 해결될

수 있을 것이다. 온라인 증명 동료평가가 적극적으로 사용되기 위해서는 자유로운 수식 입력을 가능하게 하는 것이 여전히 과제로 남아 있다.

본 연구에서는 증명 동료평가를 실시한 사례를 분석함으로써 증명 동료평가의 신뢰도와 타당도를 조사하였다. 이를 통해 앞으로의 증명 동료평가의 활용에 대한 함의를 이끌어 낸 한편, 본 연구는 증명 동료평가를 통해 학생들에게 어떠한 학습이 일어나는지 그리고 증명 동료평가의 경험이 학생들에게 어떠한 의미를 갖는지에 대해서는 설명하지 못한다는 점에서 한계를 지닌다. 교육 현장에서 증명 동료평가가 보다 적극적으로 활용되기 위해서는 증명 동료평가가 어떠한 학습효과가 있는지를 밝히는 것도 필요하다.

참고문헌

권오남 (2009. 7). 무어교수법. **대한수학회소식, 126**, 24-28.

김민정 (2008). 웹기반 형성적 동료평가 시스템 개발을 위한 설계기반 연구. **교육정보미디어연구, 14**(3), 85-114.

김정환, 조유미 (2006). 학습양식에 따라 평가유형이 수학적 성향과 문제해결력에 미치는 영향. **교육평가연구, 19**(2), 21-39.

배수정, 박주용 (2016). 대학 수업에서 누적 동료평가 점수를 활용한 성적 산출 방법의 타당성. **인지과학, 27**(2), 221-245.

신종호 (2014). 수업에 대한 형성적 동료평가 프로그램 사례 분석. **한국교원교육연구, 31**(3), 371-398.

한국교육평가학회 (2004). **교육평가 용어사전**. 서울: 학지사.

Antonini, S., & Mariotti, M. A. (2008). Indirect proof: what is specific to this way of proving?. *ZDM*, *40*(3), 401-412.

Brown, D. E., & Michel, S. (2010). Assessing proofs with rubrics: The RVF method. In *Proceedings of the 13th Annual Conference on Research in Undergraduate Mathematics Education,* Raleigh, NC. Retrieved from http://sigmaa.maa.org/rume/crume2010/Archive/Brown_D.pdf.

Bryman, A., & Cramer, D. (1990). *Quantitative data analysis for social scientists*. Taylor & Frances/Routledge.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*(3), 409-426.

Cohen, D. W. (1982). A Modified Moore Method for Teaching Undergraduate Mathematics. *The American Mathematical Monthly*, *89*(7), 473-474, 487-490.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. London; New York: Routledge.

Coppin, C. A., Mahavier, W. T., May, E. L., & Parker, E. (2009). *The*

Moore Method: A pathway to learner-centered instruction (No. 75). Washington: Mathematical Association of America.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Dochy, F., Segers, M., & Sluijsmans, D. M. A. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, *24*(3), 331-350.

Dreyfus, T. (1990). Advanced mathematical thinking. In P. Nesher & J. Kilpatrick (Eds.), *Mathematics and cognition: A research synthesis by the International Group for Psychology of Mathematics Education*(pp. 113-134). Great Britain: Cambridge University Press.

Gaillet, L. L. (1992). A Foreshadowing of Modern Theories and Practices of Collaborative Learning: The Work of Scottish Rhetorician George Jardine. Paper presented at t*he 43rd Annual Meeting of the Conference on College Composition and Communication,* Cincinnati, OH.

Gronlund, N., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. New York: Macmillan/London: Collier Macmillan.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange*, *21*, 6-13.

Harel, G., & Brown, S. (2008). Mathematical induction: Cognitive and instructional considerations. In M. Carlson & C. Rasmussen (Eds.), *Making the connection: Research and practice in undergraduate mathematics* (pp. 111-123). Washington, D. C.: Mathematical Association of America.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Issues in mathematics education* (Research in collegiate mathematics education III, Vol. 7, pp. 234-282). Providence: American Mathematics Society.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. *Second handbook of research on mathematics teaching and learning*. Greenwich: Information Age Publishing.

Hazzan, O., & Zazkis, R. (2003). Mimicry of proofs with computers: The case of linear algebra. *International Journal of Mathematical Education in Science and Technology, 34*(3), 385-402.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for research in mathematics education*, *31*(4) 396-428.

Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics, 24*(4), 389-399.

Hunter, D., & Russ, M. (1996). Peer assessment in performance studies. *British Journal of Music Education, 13*, 67-78.

Inglis, M., & Aberdein, A. (2014). Beauty is not simplicity: an analysis of mathematicians' proof appraisals. *Philosophia Mathematica*, *23*(1), 87-109.

Inglis, M., & Mejía-Ramos, J. P. (2009). The effect of authority on the persuasiveness of mathematical arguments. *Cognition and Instruction, 27*(1), 25-50.

Kim, M. (2005). *The effects of assessor's role and assessee's role on metacognitive awareness, performance, and attitude in a technology-related design task.* Unpublished doctoral dissertation, Florida State University.

Knuth, E. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education, 33*, 379-405.

Lavy, I., & Shriki, A. (2014). Engaging prospective teachers in peer assessment as both assessors and assessees: The case of geometrical proofs. *International Journal for Mathematics Teaching and Learning*. Retrieved from http://www.cimt.org.uk/journal/lavy2.pdf

Mamona-Downs, J., & Downs, M. (2005). The identity of problem solving.

The Journal of Mathematical Behavior, 24(3), 385-401.

Martin, W. G., & Harel, G. (1989). Proof frames of preservice elementary teachers. *Journal for Research in Mathematics Education, 20*(1), 41-51.

McGee, L. M., & Richgels, D. J. (1990). Learning from text using reading and writing. In T. Shanahan (Eds.), *Reading and writing together: New perspectives for the classroom* (pp. 145-169). Norwood: Christopher-Gordon Publishers, Inc.

Mejía-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*(1), 3-18.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27*(3), 249-266.

Moore, R. C. (2016). Mathematics professors' evaluation of students' proofs: A complex teaching practice. *International Journal of Research in Undergraduate Mathematics Education*, *2*(2), 246-278.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: NCTM.

Nunnaly, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

O'Donnell, A. M., & Topping, K. J. (1998). Peers assessing peers: Possibilities and problems. In K. J. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 255–278). Mahwah, NJ: Lawrence Erlbaum Associates.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 25*(1), 23-38.

Park, J. (2017). ClassPrep: A peer review system for class preparation. *British Journal of Educational Technology, 48*(2), 511-523.

Pereira-Laird, J. A., & Deane, F. P. (1997). Development and Validation of a Self-Report Measure of Reading Strategy Use. *Reading Psychology: An*

International Quarterly, 18(3), 185-235.

Pfeiffer, K. (2011). *Features and purposes of proofs in the view of novice students: Observations from proof validation and evaluation performances.* Unpublished doctoral dissertation. National University Ireland.

RAND Mathematics Study Panel. (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education.* Santa Monica: RAND.

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics, 48*(1), 83-99.

Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of extension, 37*(2), 1-5.

Schoenfeld, A. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior, 13*(1), 55-80.

Scruggs, T. E., & Mastropieri, M. A. (1998). Tutoring and students with special needs. In K. J. Topping & S. Ehly (Eds.), *Peer-assisted learning*, (pp. 165-182). Mahwahm NJ: Lawrence Erlbaum Associates.

Segal, J. (1999). Learning about mathematical proof: Conviction and validity. *The Journal of Mathematical Behavior, 18*(2), 191-210.

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for research in mathematics education*, *34*(1), 4-36.

Selden, A., & Selden, J. (2008). Overcoming students' difficulties in learning to understand and construct proofs. In M. P. Carlson & Rasmussen (Eds.), *Making the connection: Research and teaching in undergraduate mathematics education* (pp. 95-110). MAA: Washington, DC.

Selden, A., & Selden, J. (2013). *The genre of proof.* Cookeville: Tennessee Technological University.

Selden, A., & Selden, J. (2015). A comparison of proof comprehension, proof construction, proof validation and proof evaluation. *Extended Abstract for KHDM Conference*, Hanover, Germany, December 1-4, 2015.

Smith, J. C. (2006). A sense-making approach to proof: Strategies of students in traditional and problem-based number theory courses. *Journal of Mathematical Behavior, 25*, 73-90.

Smith, J. C., Nichols, S. R., Yoo, S., & Oehler, K. (2009). Building a community of inquiry in a problem-based undergraduate number theory course. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective*. New York: Routledge.

Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75.

Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (2009). *Teaching and learning proof across the grades: A K-16 perspective*. New York: Routledge.

Stylianou, D. A., Blanton, M. L., & Rotou, O. (2015). Undergraduate students' understanding of proof: Relationships between proof conceptions, beliefs, and classroom experiences with learning proof. *International Journal of Research in Undergraduate Mathematics Education, 1*(1), 91-134.

Tao, T. (2007). What is good mathematics? *Bulletin of the American Mathematical Society*, *44*(4), 623-634.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53.

Taylor, P. J. (2010). *An introduction to intraclass correlation that resolves some common confusions.* Unpublished manuscript, University of Massachusetts, Boston, USA. Retrieved from http://www.faculty.umb.edu/peter_taylor/09b.pdf.

Thompson, D. R. (1996). Learning and teaching indirect proof. *The Mathematics Teacher*, *89*(6), 474-482.

Topping, J. K. (2009). Peer assessment. *Theory Into Practice. 48*(1), 20-27.

Topping, J. K., & Ehly, S. (1998). *Peer assisted learning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Topping, J. K., Smith, F. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education, 25*(2), 149-169.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational studies in mathematics*, *48*(1), 101-119.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for research in mathematics education*, *39*(4), 431-459.

Weber, K. (2010). Mathematics Majors’ Perceptions of Conviction, Validity, and Proof. *Mathematical Thinking and Learning*, *12*, 306-336.

Weber, K. (2015). Effective proof reading strategies for comprehending mathematical proofs. *International Journal of Research in Undergraduate Mathematics Education, 1*(3), 289-314.

Weber, K. & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, *56*(2-3), 209-234.

Weber, K., Brophy, A., & Lin, K. (2008). Learning advanced mathematical concepts by reading text. In *Conference on Research in Undergraduate Mathematics Education*, San Diego, CA. Retrieved August (Vol. 7, p. 2008).

Weber, K., & Mejía-Ramos, J. P. (2014). Mathematics majors’ beliefs about proof reading. *International Journal of Mathematical Education in Science and Technology*, *45*(1), 89-103.

Yang, K. L. (2012). Structures of cognitive and metacognitive reading strategy use for reading comprehension of geometry proof.

Educational Studies in Mathematics, *80*(3), 307-326.

Yang, K. L., & Lin, F. L. (2008). A model of reading comprehension of geometry proof. *Educational Studies in Mathematics, 67*(1), 59-76.

[부록1: 주별 동료평가 과제 목록]

주차	정리 번호	정리 내용
1주차	1.12	Let a, b, c, d and n be integers with $n > 0$. If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a+c \equiv b+d \pmod{n}$.
	1.13	Let a, b, c, d and n be integers with $n > 0$. If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a-c \equiv b-d \pmod{n}$.
	1.14	Let a, b, c, d and n be integers with $n > 0$. If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $ac \equiv bd \pmod{n}$.
	1.15	Let a, b, k, and n be integers with $n > 0$ and $k > 0$. If $a \equiv b \pmod{n}$, then $ak \equiv bk \pmod{n}$.
2주차	1.18	Let a natural number n be expressed in base 10 as $n = a_k a_{k-1} \cdots a_1 a_0$. Then if $m = a_k + a_{k-1} + \cdots + a_1 + a_0$, then $n \equiv m \pmod{3}$
	1.22	Prove the existence part of the Division Algorithm. (Hint: Given nand m, how will you define q? Once you choose this q, then how is r chosen? Then show that $0 \leq r \leq n$-1.)
	1.23	Prove the uniqueness part of the Division Algorithm. (Hint: If $nq+r=nq'+r'$, then $nq-nq'=r'-r$. Use what you know about rand r' as part of your argument that $q=q'$.)
	Exercise 1.31	(Euclidean Algorithm). Using the previous theorem and the Division Algorithm successively, devise a procedure for finding the greatest common divisor of two integers.
3주차	1.34	Let a and b be integers. If $(a,b)=1$, then there exist integers x and y such that $ax+by=1$.
	1.38	Let a, b and n be integers. If $(a,n)=1$ and $(b,n)=1$, then $(ab,n)=1$.
	1.42	Given integers a, b, and c with a and b not both 0, there exist integers x and y that satisfy the equation $ax+by=c$ if and only if $(a,b)\|c$.

	1.47	Let a, b and c be integers. Then the equation $ax+by=c$ has a solution if and only if $(a,b)\|c$. If x_0, y_0 is a solution, that is, $ax_0+by_0=c$, then for every integer k, the integers $x=x_0+k\frac{b}{(a,b)}$ and $y=y_0-k\frac{a}{(a,b)}$ also satisfy the linear Diophantine equation $ax+by=c$. Moreover, every solution to the linear Diophantine equation $ax+by=c$ is of this form.
4주차	1.51	If a and b are natural numbers, then $gcd(a,b)lcm(a,b)=ab$.
	2.1	If n is a natural number greater than 1, then there exists a prime p such that $p\|n$.
	2.7	(Fundamental Theorem of Arithmetic (Existence Part)) Every natural number greater than 1 is either a prime number or it can be expressed as a finite product of prime numbers. That is, for every natural number n greater than 1, there exist distinct primes $p_1, p_2, \cdots, p_m$ and natural numbers $r_1, r_2, \cdots, r_m$ such that $n=p_1^{r_1}p_2^{r_2}\cdots p_m^{r_m}$.
	2.9	(Fundamental Theorem of Arithmetic (Uniqueness Part)) Let n be a natural number. Let $\{p_1, p_2, \cdots, p_m\}$ and $\{q_1, q_2, \cdots, q_s\}$ be sets of primes with $p_i \neq p_j$ if $i \neq j$ and $q_i \neq q_j$ if $i \neq j$. Let $\{r_1, r_2, \cdots, r_m\}$ and $\{t_1, t_2, \cdots, t_s\}$ be sets of natural numbers such that $n = p_1^{r_1}p_2^{r_2}\cdots p_m^{r_m} = q_1^{t_1}q_2^{t_2}\cdots q_s^{t_s}$. Then $m=s$ and $\{p_1, p_2, \cdots, p_m\}=\{q_1, q_2, \cdots, q_s\}$.
5주차	2.18	Given $(n+1)$ natural numbers, say $a_1, a_2, \ldots, a_{n+1}$, all less than or equal to $2n$, then there exists a pair, say a_i and a_j with $i \neq j$, such that $a_i \| a_j$.
	Exercise 2.23	Show that is $7^{\frac{1}{3}}$ irrational.

	2.27	Let p be a prime and let a and b be integers. If $p\|ab$, then $p\|a$ or $p\|b$.
	2.34	Let k be a natural number. Then there exists a prime larger than k.
6주차	2.35	(Infinitude of Prime Theorem) There are infinitely many prime numbers.
	2.38	(Infinitude of $4k+3$ Primes Theorem). There are infinitely many prime numbers that are congruent to 3 (mod 4).
	2.46	There exist arbitrarily long strings of consecutive composite numbers. That is, for any natural number n there is a string of more than n consecutive composite numbers.
	Question 3.6	(Describe technique) Let a, n, and $r$be natural numbers. Describe how to find a number k ($0 \leq k \leq n-1$) such that $k \equiv a^r (\mathrm{mod}\ n)$ subject to the restraint that you never multiply numbers larger than $n$and that you only have to do about $\log_2 r$ such multiplications.
8주차	3.13	Suppose $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ is a polynomial of degree $n>0$ with integer coefficients. Then $f(x)$ is a composite number for infinitely many integers x.
	3.17	Let n be a natural number. Any set, $\{a_1, a_2, \cdots, a_n\}$, of $n$integers for which no two are congruent modulo n is a complete residue system modulo n.
	3.28	(Chinese Remainder Theorem). Suppose $n_1, n_2, \cdots, n_L$ are positive integers that are pairwise relatively prime, that is, $(n_i, n_j)=1$ for $i \neq j$, $1 \leq i, j \leq L$. Then the system of L congruences $x \equiv a_1 (\mathrm{mod}\ n_1)$ $x \equiv a_2 (\mathrm{mod}\ n_2)$ $\vdots$ $x \equiv a_L (\mathrm{mod}\ n_L)$ has a unique solution modulo the product $n_1 n_2 \cdots n_L$.

	4.4	Let a and n be natural numbers with $(a, n) = 1$. Then there exist natural numbers i and j with $i \neq j$ such that $a^i \equiv a^j \pmod{n}$.
	4.10	Let a and n be natural numbers with $(a, n) = 1$, let $k = \mathrm{ord}_n(a)$, and let m be a natural number. Then $a^m \equiv 1 \pmod{n}$ if and only if $k \mid m$.
	4.11	Let a and n be natural numbers with $(a, n) = 1$. Then $\mathrm{ord}_n(a) < n$.
9주차	4.13	Let p be a prime and let a be an integer not divisible by p, that is, $(a, p) = 1$. Then $\{a, 2a, 3a, \cdots, pa\}$is a complete residue system mod p.
	4.31	Let n be a natural number and let $x_1, x_2, \ldots, x_{\phi(n)}$be the distinct natural numbers less than or equal to n that are relatively prime to n. Let a be a non-zero integer relatively prime to n and let i and j be different natural numbers less than or equal to $\phi(n)$. Then $ax_i \not\equiv ax_j \pmod{n}$.
	4.32	(Euler's Theorem)If a and n are integers with $n > 0$ and $(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$.
	4.36	Let p be a prime and let a be an integer with $1 \leq a < p$. Then there exists a unique natural number b less than p such that $ab \equiv 1 \pmod{p}$.
10주차	4.42	(Converse of Wilson's Theorem) If n is a natural number such that $(n-1)! \equiv -1 \pmod{n}$, then n is prime.
	5.3	Let p and q be distinct primes and E be a natural number relatively prime to $(p-1)(q-1)$. Then there are natural numbers D and y such that $$ED = 1 + y(p-1)(q-1).$$
11주차	6.3	(Lagrange's Theorem). If p is a prime and $f(x)=a_nx^n+ a_{n-1}x^{n-1}+\cdots+ a_0$ is a polynomial with integer coefficients and $a_n \not\equiv 0 \pmod{p}$, then $f(x) \equiv 0 \pmod{p}$ has at most n non-congruent solutions modulo p.
	6.4	Suppose p is a prime and $\mathrm{ord}_p(a)=d$. Then for each natural number I with $(i, d)=1$, $\mathrm{ord}_p(a^i)=d$.

	6.6	Let p be a prime and suppose g is a primitive root modulo p. Then the set $\{0, g, g^2, g^3, \cdots, g^{p-1}\}$ forms a complete residue system modulo p.
	6.13	If p and q are two different primes, then $\sum_{d\mid pq}\phi(d)=pq$.
12주차	6.17	Every prime p has $\phi(p-1)$ primitive roots.
	6.23	If n and m are relatively prime natural numbers, then $\phi(mn)=\phi(m)\phi(n)$.
	7.9	(Euler's Criterion) Suppose p is an odd prime and p does not divide the natural number a. Then a is a quadratic residue modulo p if and only if $\alpha^{\frac{p-1}{2}} \equiv 1$ (mod p) ; and a is a quadratic non-residue modulo p if and only if $\alpha^{\frac{p-1}{2}} \equiv -1$ (mod p). This criterion can be abbreviated using the Legendre symbol: $\alpha^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right)$ (mod p).
	7.12	(Infinitude of $4k+1$ Primes Theorem) There are infinitely many primes congruent to 1 modulo 4. (Hint: If $p_1, p_2, \cdots, p_r$ are primes each congruent to 1 modulo 4, what can you say what each prime factor of the number $N=(2p_1p_2...p_r)^2+1$)

[부록2: 증명 평가과제 1 - 학생용]

다음은 정리에 대한 증명들을 작성한 것입니다.

제시된 증명이 동료가 작성한 것이라 가정하고, 각 증명을 현재 Classprep 과제에서 사용하고 있는 채점기준을 적용하여 채점해주세요. 그리고 그렇게 채점한 이유를 써주세요.

<참고>는 평가할 증명에서 사용한 정리를 나타낸 것입니다. 또한 답안에는 흐름에 따라 번호가 부여 되었습니다. 평가영역에 따라 그렇게 평가한 근거가 되는 문장을 번호로 명시하고, 이유를 제시해 주시기 바랍니다.

(Infinitude of Prime Theorem) There are infinitely many prime numbers.

〈참고〉

2.34. **Theorem**. Let k be a natural numbers. Then there exists a prime larger than k.

2.38. **Theorem** (Infinitude of $4k$+3 Primes Theorem). There are infinitely many prime numbers that are congruent to 3 (mod 4).

증명 1)

귀류법을 사용하기 위해, 소수가 무한히 많다고 가정해보자. ························ ①

그러면 가장 큰 소수가 존재한다. ··· ②

*k*를 가장 큰 소수라고 하자. ··· ③

그러나 정리 2.34에 의해 *k*보다 큰 소수가 존재하는데

이것은 *k*가 가장 큰 소수라는 데에 모순이다. ······································· ④

그러므로 귀류법 가정이 틀렸다. 따라서 소수가 무한히 많음이 증명되었다. ·· ⑤

논리성: 1 / 3 / 5점

이유:

명료성: 1 / 2 / 3점

이유:

참신성: 1 / 2점
이유:

증명 2)
소수가 유한하다고 하고 차례로 $p_1, p_2, \ldots, p_n$이라고 하자. .... ①
$q = (p_1 p_2 \cdots p_n) + 1$라는 수를 생각해보자. .... ②
Fundamental Theorem of Arithematic에 의해, q는 유한한 소수들의 곱으로 표현된다. 하지만 $p_1, p_2, \cdots p_n$ 어느 것도 q를 나누지 못한다. .... ③
즉 q는 소수다. 따라서 소수의 개수는 무한하다. .... ④

논리성: 1 / 3 / 5점
이유:

명료성: 1 / 2 / 3점
이유:

참신성: 1 / 2점
이유:

증명 3)

$n \to \infty$ 일 때, $\dfrac{\pi(n)}{\dfrac{\log n}{n}} \to 1$ 이다. ·· ①

그러므로 $n \to \infty$ 일 때, $\pi(n) \to \infty$ 이다. ··· ②

논리성: 1 / 3 / 5점

이유:

명료성: 1 / 2 / 3점

이유:

참신성: 1 / 2점

이유:

증명 4)

소수의 개수가 유한하다고 가정하자. ········ ①
그러면 $4k+3$꼴의 소수도 유한하다고 할 수 있다. ········ ②
이는 $4k+3$꼴의 소수가 무한하다는 정리 2.38에 모순이다. ········ ③
따라서 소수의 개수는 무한하다. ········ ④

논리성: 1 / 3 / 5점
이유:

명료성: 1 / 2 / 3점
이유:

참신성: 1 / 2점
이유:

증명 5) 먼저 소수가 유한하다고 가정하자. ······························· ① 그러면 그 유한한 n개의 소수의 곱에 1을 더한 값을 생각하면, ········ ② n개의 소수는 모두 그 수를 나누지 못하므로 그 값도 새로운 소수가 되는데 이는 소수가 유한하다고 한 것에 모순이다. ························· ③ 따라서 소수는 무한하다고 할 수 있다. ····························· ④
논리성: 1 / 3 / 5점 이유: 명료성: 1 / 2 / 3점 이유: 참신성: 1 / 2점 이유:

채점 기준

1) 논리성 Logical Reasoning

5점	논리적인 비약이나 오류가 없으며 증명이 완성되었다.
3점	증명의 군데군데에 논리적인 비약 또는 논리적인 오류가 있다. 또는 증명이 절반 정도 밖에 완성되지 않았다.
1점	논리적인 비약이나 오류가 심하다. 또는 증명이 거의 완성되지 않았다.

2) 명료성 Clarity

3점	증명의 구성이 깔끔하다. 수학적 표현이 매끄럽다.
2점	증명에 필요 없는 내용이 조금 들어있다. 수학적 표현이 약간 매끄럽지 않다.
1점	증명에 필요 없는 내용이 많다. 수학적 표현이 매끄럽지 않다.

3) 참신성 Novelty

2점	참신하다.
1점	그저 그렇다.

[부록3: 증명 평가과제 2-전문가용]

다음은 정리에 대한 증명들을 작성한 것입니다.

제시된 증명을 학부생이 작성한 것이라 가정하고, 각 증명을 주어진 채점기준을 적용하여 채점하고, 그렇게 채점한 이유를 써주시길 바랍니다.

<참고>는 평가할 증명에서 사용한 정리를 나타낸 것입니다. 또한 답안에는 흐름에 따라 번호가 부여 되었습니다. 평가영역에 따라 그렇게 평가한 근거가 되는 문장을 번호로 명시하고, 이유를 제시해 주시기 바랍니다.

(Fermat's Little Theorem)

If p is a prime and a is an integer relatively prime to p, then $a^{(p-1)} \equiv 1 \pmod p$.

<참고>

1.38. **Theorem**. Let a, b and n be integers. If $(a,n)=1$ and $(b,n)=1$, then $(ab,n)=1$.

1.40. **Theorem**. Let a, b, c, and n be integers with $n>0$. If $ac\equiv bc$ (mod n) and $(c, n)=1$, then $a\equiv b$(mod n).

4.5. **Theorem**. Let a, b, c and n be integers with $n>0$. If $ac\equiv bc$(mod n) and $(c, n)=1$, then $a\equiv b$(mod n).

4.14. **Theorem**. Let p be a prime and let a be an integer not divisible by p. Then $a\cdot 2a\cdot 3a\cdot\cdots\cdot(p-1)a \equiv 1\cdot 2\cdot 3\cdot\cdots\cdot(p-1)$ (mod p).

4.16. **Theorem**. (Fermats Little Theorem, Version II). If p is a prime and a is an integer, then $a^p\equiv a$ (mod p).

4.25. **Lemma**. If p is prime and i is a natural number less than p, then p divides $\binom{p}{i}$.

4.32. **Theorem.** (Euler's Theorem) If a and n are integers with $n>0$ and $(a, n)=1$, then $a^{\phi(n)}\equiv 1$ (mod n).

증명 1) 정리 4.14에 의해, $a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}$이다. ································ ① p는 소수이므로 $p \nmid (p-1)!$이다. ·· ② 따라서 $(a,p)=1$이므로 정리 1.40에 의해, $a^{p-1} \equiv 1 \pmod{p}$이다. ··········· ③
논리성: 1 / 3 / 5점 이유: 명료성: 1 / 2 / 3점 이유: 참신성: 1 / 2점 이유:

증명 2)

p는 소수이고 a는 p와 서로소인 정수라고 가정하자. ······································ ①

그러면 정리 4.14에 의해 $a \cdot 2a \cdot 3a \cdots (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdots (p-1) \pmod{p}$

이고 즉, $1 \cdot 2 \cdot 3 \cdots (p-1)a^{p-1} \equiv 1 \cdot 2 \cdot 3 \cdots (p-1) \pmod{p}$이다. ··········· ②

소수의 정의에 의해 $(1,p)=(2,p)=\cdots=(p-1,p)=1$이다. ···························· ③

정리 1.38을 $(p-1)$번 적용하면, $(1 \cdot 2 \cdots (p-1) \cdot p)=1$을 얻는다. ·········· ④

그러면 정리 4.5에 의해 $a^{p-1} \equiv 1 \pmod{p}$이다. ··· ⑤

논리성: 1 / 3 / 5점

이유:

명료성: 1 / 2 / 3점

이유:

참신성: 1 / 2점

이유:

증명 3)

소수 p와 p와 서로소인 정수 a를 생각하자. ····· ①

정리 4.16에 의하여 $a^p \equiv a \pmod{p}$인데, ····· ②

$(a,p) \equiv 1$이므로 정리 4.5에 의하여 합동식의 양변을 a로 나누면 ····· ③

$a^{p-1} \equiv 1 \pmod{p}$이다. ····· ④

논리성: 1 / 3 / 5점

이유:

명료성: 1 / 2 / 3점

이유:

참신성: 1 / 2점

이유:

증명 4)

$a^p \equiv a \pmod p$를 증명한다. ······ ①

$a=1$이면, $1^{p-1} \equiv 1 \pmod p$이다. ······ ②

이제 $a \le k \in N$인 a에 대해 명제가 참이라고 가정하자. ······ ③

이항정리에 의해 $k \ge 2$는 $(k+1)^p = k^p + \binom{p}{1}k^{p-1} + \binom{p}{2}k^{p-2} + \cdots + 1$ 를 만족한다. ······ ④

그러므로 $(k+1)^p \equiv k^p + \binom{p-1}{1}k^{p-2} + \cdots + \binom{p-1}{p-2}k + 1 \pmod p$

$\equiv k + \binom{p}{1}k^{p-1} + \cdots + 1 \pmod p$이다. (귀류법가정에 의해서) ······ ⑤

정리 4.25에 의하여 $i < p$이면 $p \mid \binom{p}{i}$ 이다. ······ ⑥

귀납법에 의해, 모든 자연수 a에 대해 $a^p \equiv a \pmod p$이다. ······ ⑦

그러면 $p \mid a(a^p - 1)$이다. ······ ⑧

$(a,p)=1$이므로 $p \mid a^p - 1$이고 따라서 $a^p \equiv 1 \pmod p$이다. ······ ⑨

논리성: 1 / 3 / 5점

이유:

명료성: 1 / 2 / 3점

이유:

참신성: 1 / 2점

이유:

증명 5)

Euler's Theorem에 의해 $a^{\phi(p)} \equiv 1 \pmod{p}$가 성립한다. ································ ①
이때, p의 정의에 의해 p는 p보다 작은 어떤 자연수들과 서로소이므로 ··· ②
$\phi(p) = p-1$이 되어 ·· ③
$a^{p-1} \equiv 1 \pmod{p}$가 성립한다. ·· ④

논리성: 1 / 3 / 5점
이유:

명료성: 1 / 2 / 3점
이유:

참신성: 1 / 2점
이유:

채점 기준

1) 논리성 Logical Reasoning

5점	논리적인 비약이나 오류가 없으며 증명이 완성되었다.
3점	증명의 군데군데에 논리적인 비약 또는 논리적인 오류가 있다. 또는 증명이 절반 정도 밖에 완성되지 않았다.
1점	논리적인 비약이나 오류가 심하다. 또는 증명이 거의 완성되지 않았다.

2) 명료성 Clarity

3점	증명의 구성이 깔끔하다. 수학적 표현이 매끄럽다.
2점	증명에 필요 없는 내용이 조금 들어있다. 수학적 표현이 약간 매끄럽지 않다.
1점	증명에 필요 없는 내용이 많다. 수학적 표현이 매끄럽지 않다.

3) 참신성 Novelty

2점	참신하다.
1점	그저 그렇다.

[부록4: 주차별 학생들이 받은 동료평가 전체점수의 평균]

학생	1주차	2주차	3주차	4주차	5주차	6주차	8주차	9주차	10주차	11주차	12주차
A	8.6675	8.9175	8.5825	8.5825	8.9575	8.3325	8.8775	8.915	8.5	8.5025	8.165
B	8.71	8.9575	8.5	9.04	8.9175	8.8325	8.8325	8.665	8.7075	8.665	9.0825
C	7.96	8.2163	8.9175	9	6.915	8.3325	8.2163	8.75	8	8.165	8.2925
D	8.75	8.5017	8.4175	8.9175	8.5425	8.7475	8.5567	8.25	8.915	8.585	9.25
E	8.25	8.3317	7.8325	9	7.8325	8.665	7.245	8.7925	8.7075	8.8325	8.915
F	9	8.75	7.915	8.75	7.915	8.7525	8.21	8.42	8.8325	8.4975	9.22
G	9	8.5833	9.0825	8.5825	7.915	8.915	9.0825	8.9175	8.8118	8.8325	9.2075
H	8.9175	8.3333	8.7475	8.5825	9.0425	8.7475	7.7525	8.8325	8.2767	8.0825	9.1225
I	8.9575	9.25	9.0825	8.83	8.665	9	9	8.915	8.9975	8.9481	8.8325
J	9	8.875	9	8.835	8.915	9.165	8.7075	8.9175	9	9.0825	9.08
K	8.7925	8.3342	8.5825	8.7475	8.7475	9	8.33	8.4175	8.75	8.8867	7.125
L	9.125	8.8325	8.5	9.0825	8.4575	8.75	8.8325	8.7075	9.165	8.9175	9
M	7.7525	8.9158	8.5825	9.0825	8.75	8.25	8.96	9	8.1675	8.0825	8.9575
N	8.8325	8.375	8.9175	7.75	9.1675	9	9.1667	9	8.9575	8.6633	8.3333
O	8.7475	8.6658	9	8.5825	7.8325	8.75	7.9175	8.415	9.0425	9	8.8325
P	8.4175	8.7483	9	7.75	8.915	9.0825	9.6675	9	8.4175	8.83	9

[부록5: 주차별 학생들이 받은 동료평가 논리성 점수의 평균]

학생	1주차	2주차	3주차	4주차	5주차	6주차	8주차	9주차	10주차	11주차	12주차
A	4.6675	4.9175	4.8325	4.4175	5	4.3325	4.335	4.5	4.3325	4.585	4.1675
B	4.875	4.75	4.5	4.8325	5	4.8325	4.8325	4.665	4.5825	4.5	4.875
C	4.25	4.3644	4.75	5	3.415	4.5	4.3644	4.8325	4.085	4.3325	4.5
D	5	4.4167	4.335	4.75	4.375	4.665	4.3333	4.4175	4.75	4.5825	4.75
E	4.75	4.4992	4	5	4.0825	4.5	3.665	5	4.5825	4.8325	4.8325
F	5	4.5825	3.915	4.6675	3.915	5	4.4175	4.585	4.8325	4.415	5
G	4.9175	4.5	4.9175	4.5	4.0825	4.8325	5	5	4.75	4.75	5
H	4.75	4.25	4.5825	4.8325	5	4.665	3.9175	4.8325	4.22	4.3325	5
I	5	5	5	4.8325	4.665	5	4.8325	4.8325	4.8325	4.8883	4.8325
J	5	5	4.9175	4.75	4.8325	5	4.5	5	4.8325	5	4.8325
K	4.875	4.5	4.3325	4.415	4.6675	4.8325	4.3325	4.6675	4.6675	4.7767	3.5
L	5	4.8325	4.5	5	4.5	4.8325	4.665	4.7075	4.75	4.9175	5
M	4.335	4.8325	4.5	5	4.8325	4.4175	4.875	5	4.3325	4	4.8325
N	4.5825	4.5	5	3.8325	4.8325	5	5	5	4.8325	4.5533	4.3333
O	4.665	4.665	5	4.585	4	4.75	4	4.5825	4.875	5	4.5
P	4.5	4.9167	4.8325	4.335	4.8325	5	5	5	4.4175	4.8325	5

[부록6: 주차별 학생들이 받은 동료평가 명료성 점수의 평균]

학생	1주차	2주차	3주차	4주차	5주차	6주차	8주차	9주차	10주차	11주차	12주차
A	3	2.9175	2.5	2.8325	2.7925	2.9175	3	3	2.9175	2.9175	2.75
B	2.835	2.9167	3	3	2.71	2.835	3	3	2.9175	3	3
C	2.585	2.7689	2.9175	3	2.5	2.75	2.7689	2.9175	2.8325	2.75	2.6675
D	2.75	2.9175	2.9175	2.7925	2.835	3	2.89	2.8325	3	2.835	3
E	2.5	2.75	2.5825	3	2.75	2.9175	2.3325	2.7925	2.875	2.9175	3
F	3	3	3	2.9175	2.835	2.7525	2.625	2.835	2.9175	3	2.89
G	3	2.875	2.9175	2.9175	2.75	3	3	2.7925	2.917	2.9175	3
H	2.9175	2.8342	3	2.6675	2.7925	3	2.5425	3	2.89	2.75	2.8325
I	2.875	2.8342	3	2.8325	2.875	2.9175	3	3	3	2.90769	2.835
J	2.9175	2.875	3	2.9175	2.835	3	2.9175	2.835	3	2.75	3
K	2.9175	2.7517	3	3	2.7075	2.9175	2.8325	2.75	3	3	2.4175
L	2.875	3	2.75	3	2.875	2.835	3	3	3	3	3
M	2.4175	3	3	3	2.9175	2.75	2.9175	3	2.835	3	2.9175
N	3	2.7925	2.835	2.7925	2.835	2.89	3	3	3	3	2.6667
O	3	2.9175	2.9175	2.75	2.75	2.8325	2.7525	2.8325	3	2.9175	3
P	2.75	2.7492	2.9175	2.3325	3	3	2.875	3	2.9175	2.8325	3

[부록7: 주차별 학생들이 받은 동료평가 참신성 점수의 평균]

학생	1주차	2주차	3주차	4주차	5주차	6주차	8주차	9주차	10주차	11주차	12주차
A	1	1.0825	1.25	1.3325	1.165	1.0825	1.5425	1.415	1.25	1	1.2475
B	1	1.2908	1	1.2075	1.2075	1.165	1	1	1.2075	1.165	1.2075
C	1.125	1.0831	1.25	1	1	1.0825	1.0831	1	1.0825	1.0825	1.125
D	1	1.1675	1.165	1.375	1.3325	1.0825	1.3333	1	1.165	1.1675	1.5
E	1	1.0825	1.25	1	1	1.2475	1.2475	1	1.25	1.0825	1.0825
F	1	1.1675	1	1.165	1.165	1	1.1675	1	1.0825	1.0825	1.33
G	1.0825	1.2083	1.2475	1.165	1.0825	1.0825	1.0825	1.125	1.1448	1.165	1.2075
H	1.25	1.2492	1.165	1.0825	1.25	1.0825	1.2925	1	1.1667	1	1.29
I	1.0825	1.4158	1.0825	1.165	1.125	1.0825	1.1675	1.0825	1.165	1.1520	1.165
J	1.0825	1	1.0825	1.1675	1.2475	1.165	1.29	1.0825	1.1675	1.3325	1.2475
K	1	1.0825	1.25	1.3325	1.3725	1.25	1.165	1	1.0825	1.11	1.2075
L	1.25	1	1.25	1.0825	1.0825	1.0825	1.1675	1	1.415	1	1
M	1	1.0833	1.0825	1.0825	1	1.0825	1.1675	1	1	1.0825	1.2075
N	1.25	1.0825	1.0825	1.125	1.5	1.11	1.1667	1	1.125	1.11	1.3333
O	1.0825	1.0833	1.0825	1.2475	1.0825	1.1675	1.165	1	1.1675	1.0825	1.3325
P	1.1675	1.0825	1.25	1.0825	1.0825	1.0825	1.7925	1	1.0825	1.165	1

[부록8: 증명 동료평가 점수의 일치도]

학생	1주차				2주차				3주차			
	1.12	1.13	1.14	1.15	1.18	1.22	1.23	1.31	1.34	1.38	1.42	1.47
A						1		1	0.929			
B												
C	0.991								0.892			
D											1	0.989
E					0.989						0.908	
F												1
G					0.989						0.984	0.980
H										0.934		
I											1	
J											1	0.989
K						0.851				0.987	0.934	
L								0.947				1
M							1	0.947				0.989
N								0.991		0.984	1	
O						0.960		0.947	0.991			
P					0.989	1		0.974				1

학생	4주차				5주차				6주차		
	1.51	2.1	2.7	2.9	2.18	2.23	2.27	2.34	2.35	2.38	2.46
A		0.889				0.989					
B									0.991		
C				1		0.947			1		
D				1						0.947	
E								0.947			0.989
F									0.991		
G		0.947							1		
H		0.964					0.987				
I								0.947	1	1	0.991
J	1							0.861	1		1
K					0.892			0.989	1		
L										0.991	
M	0.989				0.991			0.885		0.947	
N		1							0.991		1
O			0.846			0.923		0.968	1		
P			0.750		1					1	

학생	8주차				9주차				10주차			
	3.13	3.17	3.28	4.4	4.10	4.11	4.13	4.31	4.32	4.36	4.42	5.3
A	0.789			0.959								
B												
C										0.500		
D				0.959								
E	1		0.934			0.991						
F	0.987			0.918			0.991					
G				1								
H	0.667								0.960			
I		1	0.987					1				
J			0.667			0.984			0.947			0.987
K	0.947		0.947						1		1	
L	0.987	0.947						1				0.989
M						1			0.960			
N												
O										0.989		
P	0.987					1						

학생	11주차				12주차			
	6.3	6.4	6.6	6.13	6.17	6.23	7.9	7.12
A					0.923			
B					0.989			
C	0.947							
D		0.960				0.987		
E		0.984		1				
F				1				
G		0.989						0.989
H	1			0.947		0.989		0.991
I								
J	0.984	0.986			0.947	0.989		0.989
K			1			0.214		
L				1				1
M		0.316		0.923	0.947			
N								0.947
O	0.984	1						0.989
P					1		1	

[부록9: 동료평가 점수 및 피드백 내용]

[12주차에 학생 K가 작성한 정리 6.23의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.214]

영역	채점자 → 피채점자		
	학생 J → 학생 K	학생 H → 학생 K	학생 D → 학생 K
논리성	1 점 - thm 6.20에 의하면 c.r.s. modulo n에서 n과 서로소인 원소들의 갯수가 pi(n)개임을 알 수 있어야하는데, 증명에서는 pi(m)에 대한 정보를 얻을 수 있다고 나온 것 같습니다. pi(m)에 대해 언급된 부분과 pi(n)에 대해 언급된 위치가 바뀌면 논리적인 증명이 될 것 같습니다.	5점 - 논리적 흐름으로는 이해가 가는 증명이었습니다.	3점 - phi(n)개의 집합은 어떻게 구성되어야하는 지 잘 이해가 안되네요.. 조금 더 자세하게 설명해주시면 이해가 잘 될 것 같습니다.
명료성	3점 - 증명의 구성이 깔끔한 것 같습니다.	2점 - 이게 당연한 내용을 증명하다 보니까 명료하게 근거를 드러내는 부분이 어려웠던 것같습니다. 행렬과 같은 배열을 활용한다면 보다 시각적으로 도움이 되지 않을까 합니다.	3점 - 필요하신 부분을 잘 설명해주신 것 같습니다.
참신성	1점 - 일반적인 증명인 것 같습니다.	1점	2점 - 새로운 방법을 시도하신 것 같아요. 설명을 꼭 들어보고 싶습니다.

[11주차에 학생M이 작성한 정리 6.4의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.316]

영역	채점자 → 피채점자		
	학생 J → 학생 M	학생 E → 학생 M	학생 D → 학생 M
논리성	5점 - 논리적인 증명인 것 같습니다.	1점 - 정리 1.40에 의해 a^k와 a^d가 합동이라고 하셨는데, a^ik와 a^id의 양변을 a^i로 나누면 그러한 합동식이 나올 수 없습니다! 양변을 1/i제곱해야만 그런 합동식이 나오는데, 그럴 수 있다는 정리를 저희가 배우진 않았군요!	1점 - 1.40은 서로소일 때 양변을 같은 수로 나눌 수 있다는 정리인데, 얻은 결과는 나눈 것이 아니라 i제곱근을 취한 결과이기 때문에 1.40에 의해 얻어지는 결과가 아닙니다. i제곱근을 취했을 때에도 합동식이 성립한다는 명제는 증명한 적이 없는 것 같습니다.
명료성	3점 - 증명에 필요한 내용만 깔끔하게 들어있는 것 같습니다.	3점 - 위 결함만 빼면 명료합니다.	3점 - 간단명료하게 서술하여 읽기에 편했습니다.
참신성	2점 - 참신한 증명인 것 같습니다 ^^	1점 - 수고하셨습니다.	1점 - 일반적인 증명인 것 같습니다.

[10주차에 학생C가 작성한 정리 4.36의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.500]

영역	채점자 → 피채점자		
	학생 K → 학생 C	학생 B → 학생 C	학생 O → 학생 C
논리성	3점 - 제 생각에는 존재성부터 보여야 하는 것 아닌 가 싶어요. 유일성은 잘 보여주셨어요.	1점 - 존재성은 보이지를 않은 것 같아요!! uniqueness part 만 보인것 같아서 너무 아쉽네요ㅠㅠ	1점 - existence도 보여야 합니다.
명료성	2점	2점 - 명료합니다! 잘 쓰셨어요!! 다음에는 존재성도!	3점 - 명료합니다.
참신성	1점	1점 - 아쉽지만 일반적인 풀이인것 같습니다 ㅠ	1점 - 무난합니다.

[8주차에 학생 H가 작성한 정리 3.13의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.667]

영역	채점자 → 피채점자		
	학생 P → 학생 H	학생 F → 학생 H	학생 L → 학생 H
논리성	2점 - a_0의 경우를 나누어 잘 풀어주셨습니다. 1. abs(a_0 >=2)인 경우 a_0*k를 대입했을 때 양수가 되게 하는 x가 무수히 많음을 언급해주셨으면 좋았을 것 같습니다. Thm 3.11로 보장되지만 언급해주셨으면 좋았을 것 같네요(사실 조건에서 a_0 > 0이 빠져있기도 합니다). 2. abs(a_0 =1)인 경우 Thm 2.46으로 연속한 세 수가 합성수가 되도록 하는 x가 많다고 하셨는데 2.46은 그 존재성을 보여줄 뿐이지, 임의의 정수 계수 n차 다항식 형태로 주어진 수가 2.46을 만족시키는지는 알 수 없겠지요? 3. a_0 =0인 경우도 고려되었으면 좋았을 것 같습니다.	3점 - a_0가 음수가 될 때는 괄호 안의 식을 -로 뒤집어 새롭게 구성해야할 것 같습니다.	3점 - a_0의 absolute value가 2 이상일때는 잘 보여 주셨습니다. 하지만 a_0가 0일 때는 빠진 것 같습니다. 또한 2.46을 이용하여 a_0의 absolute value가 1일 때를 보여주셨는데, 2.46은 임의의 n에 대해서 길이가 n인 연속적인 composite string이 존재한다는 것입니다. 이 정리를 이용해서는 a_0의 absolute value 가 1일 때 무수히 많은 x에 대해서 f(x)가 composite임을 보일 수가 없는 것 같습니다.
명료성	2점 - 풀이를 이해하기엔 문제없었지만 간단한 부분에서 언급이 조금 더 있었으면 좋았을 것 같습니다.	3점 - 경우를 나눈 것이 풀이를 간결하게 만든 것 같습니다.	3점 - 표현이 깔끔합니다.
참신성	2점 - Theorem. 2.46을 사용하려는 시도가 좋았습니다	2점 - 나름대로 참신한 풀이라고 생각합니다.	1점 - 일반적입니다.

[8주차에 학생 J가 작성한 정리 3.28의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.667]

영역	채점자 → 피채점자		
	학생 O → 학생 J	학생 G → 학생 J	학생 E → 학생 J
논리성	1점 - 죄송하지만, 잘 이해가 되지 않습니다. PMI 즉 수학적 귀납법은 모든 자연수 내지는 유한개의 자연수를 제외한 모든 자연수에 대하여 그 성질이 성립함을 보이는 방법입니다. 수학적 귀납법에 의해 마지막 부분을 증명하시면 모든 자연수에 대하여져 성질이 성립한다고 증명하신게 됩니다. 고로 이 증명은 옳지 않습니다. 그것보다는 임의의 (n_i, n_k+1)=1이므로 교재의 정리에 의해 (n_1 n_2 ... n_k, n_k+1)=1임이 성립한다. 라고 보이시는 것이 훨씬 깔끔하고 논리적입니다. 고생하셨습니다.	3점 - 좋은 증명 잘 보았습니다! 이해하기 쉽네요!!! 한 가지 여쭤보고 싶은건, 왜 굳이 수학적 귀납법을 쓰셨는지에 관한 것이에요 보통 수학적 귀납법은 무한히 많은 자연수 n에 대해 명제가 성립함을 보이려고 하는 것인데 여기서는 유한개의 자연수에 대해서만 보이면 되는거니까요. 굳이 수학적 귀납볍까지 쓸 필요가 있었을까? 하는생각이 잠깐 들어서요 수고하셨습니다!!!!!	5점 - 논리 그 자체입니다. 세세한 부분까지 증명하셨네요
명료성	3점 - 명료합니다.	3점 - 물흐르듯 깔끔한 증명이었다고 생각해요:)!!	3점 - 깔끔해요
참신성	1점 - 무난합니다.	1점 - 정석적인 풀이었다고 생각해요!!	1점 - 수고하셨습니다.

[4주차에 학생 P가 작성한 정리 2.7의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.750]

영역	채점자 → 피채점자		
	학생 K → 학생 P	학생 B → 학생 P	학생 M → 학생 P
논리성	5점 - 큰 문제는 아닌데요. 지수부분이 무한대가 되지 않는다는 것도 증명했어야 하지 않을까요?	1점 - 이 증명은 2.1 theorem 증명아닌가요? 저희가 보여야할게 n 이 소수이거나 소수의 곱으로 된 형태이다 라는 건데 지금 증명하고자하는것은 모든 n 에 대해 n을 나누는 p 가 존재한다 를 증명한거같아요	5점 - 논리적인 증명입니다.
명료성	3점 - 하지만 그 부분이 빠져서 명료하고 이해하는 데는 무척 쉬웠습니다.	1점 - 잘하신거같은데 증명이 잘못됐다고 생각합니다 제가틀린거일수도 있으니 피드백 주세요!	3점 - a와 b가 소수의 곱으로 나타내어진다는 표현보다 '소수들만의 곱'으로 나타내어진다는 말이 더 오해의 여지가 적은 것 같습니다. 잘못하면 소수 x 합성수 꼴로 나타낼 수 있다는 말로 오해할 수도 있을것 같아서요 ㅎ
참신성	1점	1점 - 안타깝지만 일반적인 풀이인거같습니다 ㅠ	1점 - 일반적인 증명입니다.

[8주차에 학생 A가 작성한 정리 3.13의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.789]

영역	채점자 → 피채점자		
	학생 J → 학생 A	학생 M → 학생 A	학생 E → 학생 A
논리성	3점 - 증명이 깔끔하고 논리적인 것 같습니다. (그냥 이건 제 개인적인 질문입니다.) f(m)이 1이 아니라면 f(m)의 절댓값이 1보다 크다라고 하셨는데, f(m)이 0이 되는 경우는 없는 건가요?	5점 - 논리적인 증명입니다.	3점 - 왜 q+q‘>1인지에 대한 조금 더 친절한 설명이 필요할 것 같습니다!
명료성	3점 - 증명이 명료하고 좋은 것 같습니다.	3점 - 명료한 증명입니다.	3점 - 간결하네요
참신성	2점 - 참신한 것 같습니다.	2점 - 귀류법을 사용하지 않고 직접 증명하신 것이 참신한 것 같습니다.	2점 - 참신해요. 수고하셨습니다.

[4주차에 학생 O가 작성한 정리 2.7의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.846]

영역	채점자 → 피채점자		
	학생 J → 학생 O	학생 I → 학생 O	학생 G → 학생 O
논리성	3점 - 증명이 잘 완성된 것 같습니다.	5점	3점 - 논리적인 증명이었던 것 같습니다! 사소한거긴 하지만 조금 신경 쓰였던 것 말씀드릴게요! 세번째 줄에 n is a prime number (itself) 라고 해주시는게 좋을 것 같네요. prime들의 곱으로 이루어진건 아니니까요. 8~9번째 줄에서 집합 S를 정의할 때에도 'Let S be the set of all natural numbers greater than 1 that are neither a prime number nor can be expressed as a finite product of prime numbers'라고 해주시는 것이 더 명확할 것 같네요. 지금처럼 정의를 해버리면 S에 2나 3같은 소수들이 포함되어버립니다. 이 부분을 고쳐야 뒤에서도 좀 더 논리적인 증명이 가능하네요. 수고하셨습니다!
명료성	3점 - 전체적인 구성이 깔끔한 것 같습니다.	3점	3점 - 증명의 흐름을 읽기가 쉬웠습니다. 감사합니다.
참신성	2점 - 증명 방법이 참신한 것 같습니다.	1점	1점 - 정석적인 풀이었다고 생각해요. 수고하셨습니다!

[2주차에 학생 K가 작성한 정리 1.22의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.851]

영역	채점자 → 피채점자		
	학생 F → 학생 K	학생 M → 학생 K	학생 H → 학생 K
논리성	3점 - 1. m-nj가 0보다 크거나 같으려면 ni 중에 m보다 작거나 같다라고 적어야 할 것 같습니다. 2. m과 n보다 작은 차이가 나는 수가 존재한다 - m-nj에 절댓값을 씌워야 이것이 정당화되지 않나 싶습니다.	5점 - 논리적인 증명입니다.	3점 - 죄송합니다만, 솔직히 증명이 잘 이해가 가지 않습니다. m보다 크고 m과 n보다 작은 차이가 나는 수를 nj로 가정하였는데, 어떻게 m-nj가 0보다 같거나 큰 게 되는 건지와 m-n이 nj보다 작게 되는 건지도 모르겠습니다. Well-Ordering Axiom을 직접적으로 가져와서 증명을 진행하는 것이 더 옳지 않을까 생각합니다.
명료성	3점 - 군더더기 없이 명료합니다.	3점 - 깔끔한 증명입니다.	2점
참신성	2점 - 다른 분들의 풀이와 다르게 ni라는 집합을 생각한 것이 나름대로 참신한 것 같습니다.	1점	1점

[5주차에 학생 J가 작성한 정리 2.34의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.861]

영역	채점자 → 피채점자		
	학생 L → 학생 J	학생 O → 학생 J	학생 I → 학생 J
논리성	5점 - 논리적입니다.	3점 - 항상 잘 보고 있습니다. 그런데 이번에 약간의 문제가 있는 거 같습니다. k이하의 모든 소수를 모았다고 해서 k보다 큰 소수가 없다. 그래서 모든 소수들을 모아놓은 집합이 된다고 한 부분이 논리적으로 오류가 있습니다.	5점
명료성	2점 - (모든 p를 곱한것 +1)은 임의의 pi로 나눠도 나머지가 생기게되서 사실 소수이고 여기서 바로 모순을 이끌어낼 수 있습니다. 그래서 굳이 pn보다 크므로 합성수이다.라고 가정을 하지않았다면 증명이 훨씬 간결해졌을 것 같습니다.	3점 - 명료합니다.	3점
참신성	1점 - 일반적입니다.	2점 - 참신합니다!! 아주 잘하셨습니다.	1점

[5주차에 학생 M이 작성한 정리 2.34의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.885]

영역	채점자 → 피채점자		
	학생 O → 학생 M	학생 J → 학생 M	학생 I → 학생 M
논리성	3점 - 우선 이 문제는 정리를 적용하기 위해서, 또한 논리를 논리적으로 전개하기 위해서 k=1일 때와 그보다 클 때로 반드시 나눠서 생각해야 합니다. (k=1이면 p=1이 나옵니다) p는 p1으로 나눠지지 않는다를 보이는 것이 핵심인데 그 부분에 있어서 너무 당연하게 넘어가셨습니다. 다른 학생들의 풀이를 보면 그 부분에 굉장히 공을 많이 들였습니다.	5점 - 증명이 잘 완성된 것 같습니다.	5점
명료성	3점 - 명료합니다.	2점 - 증명의 전체적인 구성이 깔끔한 것 같습니다.	3점
참신성	1점 - 고생하셨습니다.	2점 - Thm 2.1을 사용한 점이 참신한 것 같습니다.	1점

[4주차에 학생 A가 작성한 정리 2.1의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.889]

영역	채점자 → 피채점자		
	학생 J → 학생 A	학생 F → 학생 A	학생 P → 학생 A
논리성	3점 - S를 정의할 때, 2보다 큰 약수가 아니고 2 이상의 약수가 되어야할 것 같습니다. 이외에는 논리적 비약 없이 증명이 완성된 것 같습니다.	4점 - 다른 부분은 논리적인 오류가 없으나 첫 줄에서 '2보다 큰' 이 아니라 '2 이상' 으로 바꾸는 것이 맞는 것 같습니다. (4)	5점 - 논리적으로 완벽합니다. 접근 방법도 훌륭합니다.
명료성	3점 - 증명의 구성이 깔끔한 것 같습니다.	3점 - Well ordering Axiom 을 이용하여 깔끔하게 풀이해주신 것 같습니다. (3)	3점 - 수학적으로 필요한 부분만 잘 들어갔습니다.
참신성	2점 - 약수를 이용한 점이 참신한 것 같습니다.	2점 - 집합과 WOA를 사용하신 것이 참신하다고 생각합니다. (2)	2점 - Well-ordering axiom을 쓰는 방법은 정말 새로운 것 같습니다^^

[3주차에 학생 C가 작성한 정리 1.34의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.892]

영역	채점자 → 피채점자		
	학생 B → 학생 C	학생 F → 학생 C	학생 O → 학생 C
논리성	5점 - 아주 논리적입니다.	5점 - 처음 집합 S를 정의할 때 x가 빠진 것 빼고는 논리적인 결함이 없습니다.	3점 - gcd(a,b)=1을 let이 아니라 suppose로 처리하셔야 할 거 같습니다. 그리고 실수하신 거 같은데 S 정의하실 때 x, y in Z라 가정해야 하는 거 아닌가요?
명료성	3점 - 명료합니다.	3점 - 줄글로 설명한 부분이 읽을 때 약간 방해되는 부분이 있긴 하지만 명료합니다.	3점 - 명료합니다.
참신성	2점 - 참신합니다.	2점 - 귀류법을 사용하는 방법이 상당히 참신했던 것 같습니다.	2점 - 참신합니다.

[5주차에 학생 K가 작성한 정리 2.18의 논증에 대해 받은 동료평가 - 채점자간 일치도 0.892]

영역	채점자 → 피채점자		
	학생 B → 학생 K	학생 O → 학생 K	학생 C → 학생 K
논리성	3점 - A_k+2에서 원소를 n개 뽑으면 A_1,...A_K+1까지 원소를 뽑을 수 없다는 말은 논리적인 것 같습니다. 그런데 제가 수학을 잘 못해서 그런거 일 수도 있지만 A_k+2에서 원소를 n개 뽑지 않을 경우에는 어떻게 되는지에 대해서 언급이 없어서 좀 이해하기 어려웠던 것 같습니다. A_k+2에서 n 개를 뽑는게 유리하다? 라는 말이 왜 유리한지에 대해서 논리적인 설명이 더 필요하지 않을까 생각합니다. r_i 원소는 무엇을 얘기하는지 잘 이해가 안됩니다..ㅠㅠㅠ 죄송합니다 ㅠㅠ	3점 - 논리정연하게 잘 써주신 거 같습니다. 하지만 한 부분이 이상했습니다. 적어도 "k+2-i"개의 배수가 아니라 "k+2-r'개의 배수가 아니었나 싶습니다. 일단 점수는 잘 드렸습니다.	5점 - 본 풀이 중에 제일 와닿네요.
명료성	1점 - 뽑는다는 개념으로 풀이를 적어주셨는데 하는 등 이라는 추정의 방법은 증명에서 명료하지 못한것 같다고 개인적으로 생각합니다. 그래도 어려운 문제이니만큼 좋은 아이디어여서 좋았습니다!	3점 - 잘하셨습니다.	3점
참신성	1점 - 안타깝게도 참신한 풀이는 아니었던것 같습니다 ㅠ 화이팅!	1점 - 무난합니다.	2점

[2주차에 학생 A가 작성한 정리 1.22의 논증에 대해 받은 동료평가 - 채점자간 일치도 1]

영역	채점자 → 피채점자		
	학생 M → 학생 A	학생 K → 학생 A	학생 O → 학생 A
논리성	5점 - 논리적인 증명입니다.	5점 - S가 공집합이 아니라고 설명하는 것부터 나머지가 n 보다 작다는 것까지 논리적인 증명이었습니다.	5점 - 잘하셧습니다
명료성	3점 - 깔끔한 증명입니다.	3점 - 짧고 명료한 증명이었습니다. 좋습니다.	3점 - 잘하셨습니다
참신성	1점 - 일반적인 증명입니다.	1점	1점 - 무난한거 같습니다

[5주차에 학생 P가 작성한 정리 2.18의 논증에 대해 받은 동료평가 - 채점자간 일치도 1]

영역	채점자 → 피채점자		
	학생 M → 학생 P	학생 O → 학생 P	학생 K → 학생 P
논리성	5점 - 논리적인 증명입니다.	5점 - 잘하셨습니다.	5점 - 증명이 깔끔하고 논리적입니다. 한 집합에 속하면 그냥 divisible 하다 라고 쓸수도 있는데 예리하게 그부분까지 증명해주셔서 좋았습니다.
명료성	3점 - 명료한 증명입니다.	3점 - 잘하셨습니다.	3점
참신성	1점 - 일반적인 증명입니다.	1점 - 무난합니다.	1점

[8주차에 학생 G가 작성한 정리 4.4의 논증에 대해 받은 동료평가 - 채점자간 일치도 1]

영역	채점자 → 피채점자		
	학생 O → 학생 G	학생 D → 학생 G	학생 F → 학생 G
논리성	5점 - 완벽합니다.	5점 - 저와는 다르게 0부터 n-1까지의 표준잉여계를 이용하여 비둘기집의 원리를 명확하게 설명하신 점이 인상적이었습니다. 많이 배워갑니다.	5점 - 논리적인 오류가 없습니다. 다만 비둘기집의 원리를 이용할 때 굳이 귀류법 가정을 하지 않으셔도 될 듯 합니다.
명료성	3점 - 고생하셨습니다.	3점 - 쉽고 확실하게 이해되었습니다.	3점 - 비둘기집의 원리 적용 과정이 명료하게 잘 드러난 것 같습니다.
참신성	1점 - 무난합니다.	1점 - 일반적인 증명인 것 같습니다.	1점 - 많은 분들께서 비둘기집의 원리를 사용하셨습니다.

[12주차에 학생 L이 작성한 정리 7.12의 논증에 대해 받은 동료평가 - 채점자간 일치도 1]

영역	채점자 → 피채점자		
	학생 G → 학생 L	학생 H → 학생 L	학생 N → 학생 L
논리성	5점 - N=4k+3 꼴이 아니므로 N은 4k+3 꼴의 소수도 아니고, 가정에서 N은 4k+1꼴의 소수도 아니므로 N은 composite이라고 해주시는게 좀 더 좋을 것 같아요! 바로 N이 composite이고 odd prime factor를 가진다고해서 처음에는 이해가 잘 되지 않았어요:)	5점 - 논리적으로 흠이 없는 증명이었습니다.	5점
명료성	3점 - 깔끔한 풀이었습니다!	3점 - 간단명료하였습니다.	3점
참신성	1점 - 정석적인 풀이었습니다	1점 -	1점

[부록10: 전문가1의 증명 평가과제 1 응답 내용]

증명 1) 귀류법을 사용하기 위해, 소수가 무한히 많다고 가정해보자. ········ ① 그러면 가장 큰 소수가 존재한다. ········ ② k 를 가장 큰 소수라고 하자. ········ ③ 그러나 정리 2.34에 의해 k 보다 큰 소수가 존재하는데 이것은 k 가 가장 큰 소수라는 데에 모순이다. ········ ④ 그러므로 귀류법 가정이 틀렸다. 따라서 소수가 무한히 많음이 증명되었다. ········ ⑤
논리성: 1 / **3** / 5점 이유: 1. 단계 ①에서 유도되는 내용은 유한의 정의로부터 나와야 하므로 ②가 나오는 것은 자연스럽지 않습니다. 2. ④에서 k가 가장 큰 소수라는 데에 모순이라는 말은 다소 모호한 것 같습니다. k의 선택, 혹은 정의에 모순이라는 말이 ③과 호응되는 표현이라고 생각합니다. 3. ①의 서술을 유한개라고 가정해보자 로 해석했습니다. 오타가 아닌 경우 가정이 잘못되었으므로 논리성은 1점을 부여하겠습니다. 명료성: **1** / 2 / 3점 이유: 1. ①, ②, 그리고 ③은 소수들을 문자로 표현한다면 보다 간결하고 명확히 표현될 수 있다고 생각합니다. 2. ①에서 귀류법을 언급했으므로 ⑤는 의미가 없습니다. 참신성: **1** / 2점 이유: 증명을 위해 사용한 정리와 증명 대상이 동치입니다.
증명 2) 소수가 유한하다고 하고 차례로 $p_1, p_2, \ldots, p_n$이라고 하자. ········ ① $q = (p_1 p_2 \cdots p_n) + 1$라는 수를 생각해보자. ········ ② Fundamental Theorem of Arithematic에 의해, q는 유한한 소수들의 곱으로 표현된다. 하지만 $p_1, p_2, \cdots p_n$ 어느 것도 q를 나누지 못한다. ········ ③ 즉 q는 소수다. 따라서 소수의 개수는 무한하다. ········ ④

논리성: 1 / **3** / 5점
이유:
1. ④의 결론을 위해서는 q가 주어진 소수들과 모두 다르다는 것을 미리 언급해야 할 것입니다. 또한 귀류법을 사용하고 있으므로 q가 소수인 것으로부터 모순을 이끌어내야 합니다.

명료성: 1 / **2** / 3점
이유: 1. ①의 표현 `소수가 유한하다'는 잘못되었습니다.

참신성: 1 / **2**점
이유: 증명을 위해 사용한 정리와 증명 대상이 동치관계가 아닙니다.

증명 3)
$n \to \infty$ 일 때, $\dfrac{\pi(n)}{\dfrac{\log n}{n}} \to 1$ 이다. ········· ①
그러므로 $n \to \infty$ 일 때, $\pi(n) \to \infty$ 이다. ········· ②

논리성: **1** / 3 / 5점
이유: ①의 정리는 대수와 해석적 개념이 혼합되어 있는 정리입니다. 정수론에서는 log와 limit개념이 정의되지 않고, 따라서 위 정리를 증명할 수 없습니다. 또한 ①에서 ②를 이끌어내는 것은 $\dfrac{\log n}{n}$의 극한값이 무한이라는 풀이에서 드러나지 않은 사실로부터 뒷받침되는 것으로, 이 풀이는 직관에 의존하고 있는 것으로 보입니다.

명료성: 1 / **2** / 3점
이유:
1. 지나치게 명료성만을 추구한 나머지, 증명에 포함되어야 할 내용이 다수 빠져있습니다.
2. 정의되지 않은 기호들을 사용했습니다.

참신성: **1** / 2점
이유: 위 증명에서 이용한 정리는 소수의 무한성과 밀도를 밝히고 있어 보이고자 하는 대상보다 더 강력합니다.

증명 4) 소수의 개수가 유한하다고 가정하자. ········ ① 그러면 $4k+3$꼴의 소수도 유한하다고 할 수 있다. ········ ② 이는 $4k+3$꼴의 소수가 무한하다는 정리 2.38에 모순이다. ········ ③ 따라서 소수의 개수는 무한하다. ········ ④
논리성: 1 / 3 / **5**점 이유: 논리 전개가 자연스럽습니다. 명료성: 1 / **2** / 3점 이유: ②와 ③을 분리할 이유가 없습니다. 참신성: **1** / 2점 이유: 위 증명에서 이용한 정리는 소수의 무한성 및 특정 형태의 소수 역시 무한함을 밝히고 있어 보이고자 하는 대상보다 더 강력합니다.
증명 5) 먼저 소수가 유한하다고 가정하자. ········ ① 그러면 그 유한한 n개의 소수의 곱에 1을 더한 값을 생각하면, ········ ② n개의 소수는 모두 그 수를 나누지 못하므로 그 값도 새로운 소수가 되는데 이는 소수가 유한하다고 한 것에 모순이다. ········ ③ 따라서 소수는 무한하다고 할 수 있다. ········ ④
논리성: **1** / 3 / 5점 이유: 1. ①, ④의 표현 \`소수가 유한하다'와 \`소수가 무한하다'는 잘못되었습니다. 2. ②에서 나타난 n이 정의되지 않았습니다. 3. ③에서 새로 정의된 수가 소수임을 주장하는 과정에서, 소수의 정의를 이용하여 보이지 않았습니다. 명료성: 1 / **2** / 3점 이유: 문자를 사용하지 않은 이유로, ③에서 같은 대상을 지칭하는 표현이 다릅니다. 참신성: 1 / **2**점 이유: 다른 정리를 사용하지 않고 정의를 사용하여 잘 설계했습니다.

[부록11: 전문가2의 증명 평가과제 1 응답 내용]

증명 1) 귀류법을 사용하기 위해, 소수가 무한히 많다고 가정해보자. ······ ① 그러면 가장 큰 소수가 존재한다. ······ ② *k* 를 가장 큰 소수라고 하자. ······ ③ 그러나 정리 2.34에 의해 *k* 보다 큰 소수가 존재하는데 이것은 *k* 가 가장 큰 소수라는 데에 모순이다. ······ ④ 그러므로 귀류법 가정이 틀렸다. 따라서 소수가 무한히 많음이 증명되었다. ······ ⑤
논리성: 1 / 3 / 5점 이유: **3점**. 지금은 소수가 무한하다는 것을 증명하는 상황입니다. 그러므로 귀류법을 사용하는 경우에는 소수의 개수가 유한하다는 가정이 필요합니다. 1번 문장의 가정을 수정해야할 것 같습니다. 2번 문장에서도 '유한한 자연수 집합에서 항상 maximal element가 있다.' 는 점도 간단히 설명해주는게 좋을 것 같습니다. 명료성: 1 / 2 / 3점 이유: **2점**. 2번에서 3번 문장을 굳이 나눠서 쓸 필요가 없을 것 같습니다. '그러면 가장 큰 소수 k가 존재한다.' 는 표현이 조금 더 매끄러운 것 같습니다. 참신성: 1 / 2점 이유: **1점**. 기존에 있던 유클리드의 증명과 같이 귀류법을 사용했다는 점에서 새로울 것이 없습니다.
증명 2) 소수가 유한하다고 하고 차례로 $p_1, p_2, \ldots, p_n$이라고 하자. ······ ① $q = (p_1 p_2 \cdots p_n) + 1$라는 수를 생각해보자. ······ ② Fundamental Theorem of Arithematic에 의해, q는 유한한 소수들의 곱으로 표현된다. 하지만 $p_1, p_2, \cdots p_n$ 어느 것도 q를 나누지 못한다. ······ ③ 즉 q는 소수다. 따라서 소수의 개수는 무한하다. ······ ④
논리성: 1 / 3 / 5점 이유: **5점**. 전체적으로 괜찮은 증명인 것 같습니다. 다만 마지막 4번 문장에서 q가 p_1, ... p_n 까지의 수들 중 어떠한 것과도 같을 수 없다는 것을 설명해주면 더 좋을 것 같습니다.

명료성: 1 / 2 / 3점
이유: **3점**. 증명의 구성이나 수학적 표현이 매끄러운 것 같습니다. 마지막 4번 문장에서 ‘처음 가정과 모순이다’ 정도만 추가해주면 좋을 것 같습니다.

참신성: 1 / 2점
이유: **1점**. 기존에 잘 알려진 유클리드의 증명과 크게 다를 것이 없습니다.

증명 3)

$n \to \infty$ 일 때, $\dfrac{\pi(n)}{\dfrac{\log n}{n}} \to 1$ 이다. ······ ①

그러므로 $n \to \infty$ 일 때, $\pi(n) \to \infty$ 이다. ······ ②

논리성: 1 / 3 / 5점
이유: **1점**. 어떤 정리에 의해 1번 문장이 성립하는지에 대한 설명이 전혀 없습니다. 그리고 1번이 성립할 때 2번이 성립하는 것은 그렇게 자명해보이지 않습니다. 여기에 구체적인 증명이 필요할 것 같습니다.

명료성: 1 / 2 / 3점
이유: **1점**. 일단 $\pi(n)$이 어떤 함수인지에 그리고 n이 자연수라는 사실이 정확히 명시되어야합니다. 그리고 오른쪽 화살표를 쓰는 방법보다는 정확히 limit을 이용하여 적어주는 것이 나아보입니다.

참신성: 1 / 2점
이유: **2점**. 귀류법을 이용한 기존의 방법과는 다릅니다. 새로운 시도를 했다는 점에서 높이 평가합니다.

증명 4)

소수의 개수가 유한하다고 가정하자. ······ ①

그러면 $4k+3$꼴의 소수도 유한하다고 할 수 있다. ······ ②

이는 $4k+3$꼴의 소수가 무한하다는 정리 2.38에 모순이다. ······ ③

따라서 소수의 개수는 무한하다. ······ ④

논리성: 1 / 3 / 5점
이유: **5점**. 논리적인 비약이나 오류가 전혀 없는 것 같습니다.

명료성: 1 / 2 / 3점 이유: **2점**. 이 증명에서는 굳이 귀류법을 사용할 필요가 없습니다. '4k+3 꼴의 소수는 전체 소수 집합의 부분집합이다. 정리 2.38에 의해 이 부분집합이 무한하므로 소수의 집합도 무한하다.' 와 같은 논리가 더 좋아 보입니다. 참신성: 1 / 2점 이유: **1점**. '소수의 개수가 무한하다'를 증명하기 위해 이보다 더 강력한 정리를 사용하였으므로 별로 참신하다고 볼 수 없습니다.
증명 5) 먼저 소수가 유한하다고 가정하자. ………………………… ① 그러면 그 유한한 n개의 소수의 곱에 1을 더한 값을 생각하면, ………………… ② n개의 소수는 모두 그 수를 나누지 못하므로 그 값도 새로운 소수가 되는데 이는 소수가 유한하다고 한 것에 모순이다. ………………………… ③ 따라서 소수는 무한하다고 할 수 있다. ………………………… ④
논리성: 1 / 3 / 5점 이유: **3점**. 3번 문장에 보충해야할 부분이 필요합니다. n개의 소수가 그 수를 나누지 못할 때 그 수가 왜 새로운 소수가 되는지에 대해 설명해야합니다. 이 부분은 귀류법 가정이 모순인데 핵심 부분이라고 생각합니다. 명료성: 1 / 2 / 3점 이유: **2점**. 증명을 서술할 때 적절한 기호를 사용하지 않으면 혼란스러울 수 있습니다. 그 수라는 표현보다는 증명 2) 에서와 같은 표현이 더 명확합니다. 참신성: 1 / 2점 이유: **1점**. 기존에 있던 유클리드의 증명과 전혀 다를 것이 없습니다.

[부록12: 전문가1의 증명 평가과제 2 응답 내용]

증명 1) 정리 4.14에 의해, $a^{p-1}(p-1)! \equiv (p-1)!\ (\mathrm{mod}\ p)$이다. ············ ① p는 소수이므로 $p \nmid (p-1)!$이다. ············ ② 따라서 $(a,p)=1$이므로 정리 1.40에 의해, $a^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ············ ③
논리성: **1** / 3 / 5점 이유: 1. 문제의 조건 $(a,p)=1$과 ①에서 정리 4.14를 적용하기 위해 이용한 a가 p의 배수가 아니다는 동치가 아닙니다. 2. ②의 결론은 소수의 정의로부터 나온 것이 아니므로 엄밀성이 떨어집니다. 3. ③은 ②의 결론으로부터 $((p-1)!,p)=1$을 유도한 후 정리 1.40을 적용시켜야 합니다. 명료성: 1 / **2** / 3점 이유: 1. 지나치게 명료성만을 추구한 나머지, 증명에 포함되어야 할 내용이 다수 빠져있습니다. 2. ③의 문장 시작은 ②로부터 $(a,p)=1$를 이끌어낸 구조입니다. 참신성: 1 / **2**점 이유: 보이려는 대상보다 강력한 정리를 사용하지 않고 잘 증명하였습니다.
증명 2) p는 소수이고 a는 p와 서로소인 정수라고 가정하자. ············ ① 그러면 정리 4.14에 의해 $a \cdot 2a \cdot 3a \cdots (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdots (p-1)\ (\mathrm{mod}\ p)$ 이고 즉, $1 \cdot 2 \cdot 3 \cdots (p-1)a^{p-1} \equiv 1 \cdot 2 \cdot 3 \cdots (p-1)\ (\mathrm{mod}\ p)$이다. ············ ② 소수의 정의에 의해 $(1,p)=(2,p)=\cdots=(p-1,p)=1$이다. ············ ③ 정리 1.38을 $(p-1)$번 적용하면, $(1 \cdot 2 \cdots (p-1) \cdot p)=1$을 얻는다. ············ ④ 그러면 정리 1.40에 의해 $a^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ············ ⑤
논리성: 1 / **3** / 5점 이유: 1. 문제의 조건 $(a,p)=1$과 ②에서 정리 4.14를 적용하기 위해 이용한 a가 p의 배수가 아니다는 동치가 아닙니다.

2. ③은 소수의 정의로부터 나온 성질일 뿐, 정의가 아닙니다. 명료성: 1 / **2** / 3점 이유: 1. ②와 ⑤사이에 여러 식이 존재함에도 불구하고, ⑤에서 정리 1.40을 어느 식에 적용하는지가 드러나 있지 않습니다. 2. ④의 표기가 잘못되었습니다. 특별한 감점 요소는 아닌 것 같습니다. 3. If~, then~ 형태의 정리는 If의 가정을 풀이에서 다시 가정하는 경향이 있는데, 이것은 개인 취향에 가깝다고 생각하기 때문에 특별히 언급하지 않겠습니다. 참신성: 1 / **2**점 이유: 보이려는 대상보다 강력한 정리를 사용하지 않고 잘 증명하였습니다.
증명 3) 소수 p와 p와 서로소인 정수 a를 생각하자. ······ ① 정리 4.16에 의하여 $a^p \equiv a \pmod{p}$인데, ······ ② $(a,p) \equiv 1$이므로 정리 4.5에 의하여 합동식의 양변을 a로 나누면 ······ ③ $a^{p-1} \equiv 1 \pmod{p}$이다. ······ ④
논리성: 1 / 3 / **5**점 이유: 논리적으로 문제가 없는 증명이라 생각합니다. 명료성: 1 / **2** / 3점 이유: 1. ③의 '양변을 a로 나누면' 이라는 표현은 잘못되었습니다. Cancellation 개념은 아직 배우지 않았겠지만, 크게 잘못된 표현이기에 감점하였습니다. 참신성: **1** / 2점 이유: 정리 4.16이 보다 일반적인 내용을 다루고 있다고 생각합니다.
증명 4) $a^p \equiv a \pmod{p}$를 증명한다. ······ ① $a=1$이면, $1^{p-1} \equiv 1 \pmod{p}$이다. ······ ② 이제 $a \leq k \in N$인 a에 대해 명제가 참이라고 가정하자. ······ ③ 이항정리에 의해 $k \geq 2$는 $(k+1)^p = k^p + \binom{p}{1}k^{p-1} + \binom{p}{2}k^{p-2} + \cdots + 1$ 를 만족한다.

·· ④

그러므로 $(k+1)^p \equiv k^p + \binom{p-1}{1}k^{p-2} + \cdots + \binom{p-1}{p-2}k + 1 \pmod{p}$

$\equiv k + \binom{p}{1}k^{p-1} + \cdots + 1 \pmod{p}$이다. (귀류법가정에 의해서) ··································· ⑤

정리 4.25에 의하여 $i < p$이면 $p \mid \binom{p}{i}$ 이다. ··· ⑥

귀납법에 의해, 모든 자연수 a에 대해 $a^p \equiv a \pmod{p}$이다. ··································· ⑦

그러면 $p \mid a(a^p - 1)$이다. ·· ⑧

$(a,p) = 1$이므로 $p \mid a^p - 1$이고 따라서 $a^p \equiv 1 \pmod{p}$이다. ·································· ⑨

논리성: **1** / 3 / 5점

이유:

1. ⑤에서 근거로 세운 귀류법 가정은 언급된 바 없습니다. 귀납법 가정이라 할지라도 이를 어떻게 적용하겠다는 언급 없이, k^p에는 이를 적용하고 이전 식에는 없는 $\binom{p}{1}k^{p-1}$항이 등장시켜 이에는 적용하지 않는 것은 오류라고 생각합니다.
2. ⑦의 결론을 내리기 위해서 필요한 $(k+1)^p \equiv k+1 \pmod{p}$가 어떻게 유도되는지에 대한 설명이 없습니다.
3. ⑧부터 a 지수가 틀렸습니다. 이로 인해 결과 역시 잘못되었습니다.
4. ⑨에서 $(a,p) = 1$로부터 $p \mid a^p - 1$를 이끌어내기 위해서는 다른 정리가 필요합니다.

명료성: **1** / 2 / 3점

이유:

1. ①은 필요없는 항목입니다.
2. ③에서 k에 대한 설명이 없습니다.
3. ④의 문장은 $k \geq 2$가 등식이 성립하는 필요조건처럼 기술되어 있습니다.
4. ⑤의 두 번째 식에서 임의로 식을 생략하여 혼돈을 주고, 세 번째 식은 일반항을 알아볼 수 없습니다.
5. ⑥의 i에 대한 조건이 없습니다.

참신성: 1 / **2**점

이유: 이항정리를 활용한 귀납법에 의한 풀이가 독창적이라고 생각합니다.

증명 5) Euler's Theorem에 의해 $a^{\phi(p)} \equiv 1 \pmod p$가 성립한다. .... ① 이때, p의 정의에 의해 p는 p보다 작은 어떤 자연수들과 서로소이므로 .... ② $\phi(p) = p-1$이 되어 .... ③ $a^{p-1} \equiv 1 \pmod p$가 성립한다. .... ④
논리성: 1 / 3 / **5**점 이유: 1. ②는 소수의 정의가 아닌 성질입니다. 명료성: 1 / **2** / 3점 이유: 1. ②의 결론은 p의 정의가 아니라 prime의 정의로부터 나옵니다. 또한 어떤 자연수가 아닌 모든 자연수입니다. 참신성: **1** / 2점 이유: Euler's Theorem은 보이고자 하는 정리의 일반화입니다.

[부록13: 전문가2의 증명 평가과제 2 응답 내용]

증명 1) 정리 4.14에 의해, $a^{p-1}(p-1)! \equiv (p-1)!\ (\mathrm{mod}\ p)$이다. ············ ① p는 소수이므로 $p \nmid (p-1)!$이다. ············ ② 따라서 $(a,p)=1$이므로 정리 1.40에 의해, $a^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ············ ③
논리성: 1 / 3 / 5점 이유: **3점**. 2번에서 p가 (p-1)!을 나누지 못하는 이유를 언급해줄 필요가 있을 것 같습니다. 명료성: 1 / 2 / 3점 이유: **3점**. 증명의 구성이 매끄럽고 불필요한 표현은 없는 것 같습니다. 참신성: 1 / 2점 이유: **1점**. 일반 정수론 교재에서 보이는 가장 흔한 방법이므로 참신하다고 볼 수 없습니다.
증명 2) p는 소수이고 a는 p와 서로소인 정수라고 가정하자. ············ ① 그러면 정리 4.14에 의해 $a \cdot 2a \cdot 3a \cdots (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdots (p-1)\ (\mathrm{mod}\ p)$이고 즉, $1 \cdot 2 \cdot 3 \cdots (p-1)a^{p-1} \equiv 1 \cdot 2 \cdot 3 \cdots (p-1)\ (\mathrm{mod}\ p)$이다. ············ ② 소수의 정의에 의해 $(1,p)=(2,p)=\cdots=(p-1,p)=1$이다. ············ ③ 정리 1.38을 $(p-1)$번 적용하면, $(1 \cdot 2 \cdots (p-1) \cdot p)=1$을 얻는다. ············ ④ 그러면 정리 4.5에 의해 $a^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ············ ⑤
논리성: 1 / 3 / 5점 이유: **5점**. 각 문장 사이에 논리적인 비약이나 오류는 보이지 않습니다. 논리 전개에서 수정할 부분은 없는 것 같습니다. 명료성: 1 / 2 / 3점 이유: **3점**. 문제에서 이미 p와 a에 관한 언급이 있었는데 1번 문장에서 이 사실을 또 쓸 필요는 없습니다. 이 부분만 제외하면 다른 표현에서 문제는 없는 것 같습니다. 참신성: 1 / 2점 이유: **1점**. 증명 1)과 더불어서 페르마 정리를 증명하는데 이용되고 있는 가장 흔한 방법인 것 같습니다.

증명 3) 소수 p와 p와 서로소인 정수 a를 생각하자. ······ ① 정리 4.16에 의하여 $a^p \equiv a\ (\mathrm{mod}\ p)$인데, ······ ② $(a,p) \equiv 1$이므로 정리 4.5에 의하여 합동식의 양변을 a로 나누면 ······ ③ $a^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ······ ④
논리성: 1 / 3 / 5점 이유: **5점**. 각 문장사이에 논리적인 비약이나 오류는 전혀 보이지 않습니다. 명료성: 1 / 2 / 3점 이유: **3점**. 문제에서 이미 p와 a가 어떤 수인지에 대해서 언급이 있었는데 1번 문장을 다시 적어줄 필요는 없었던 것 같습니다. 전체적인 흐름은 괜찮은 것 같습니다. 참신성: 1 / 2점 이유: **1점**. 이 증명은 기초정수론에서 전형적으로 사용하는 방법을 나열했을 뿐입니다. 특별히 뛰어나다고 할 만한 아이디어가 사용되지 않은 것 같습니다.
증명 4) $a^p \equiv a\ (\mathrm{mod}\ p)$를 증명한다. ······ ① $a=1$이면, $1^{p-1} \equiv 1\ (\mathrm{mod}\ p)$이다. ······ ② 이제 $a \leq k \in N$인 a에 대해 명제가 참이라고 가정하자. ······ ③ 이항정리에 의해 $k \geq 2$는 $(k+1)^p = k^p + \binom{p}{1}k^{p-1} + \binom{p}{2}k^{p-2} + \cdots + 1$ 를 만족한다. ······ ④ 그러므로 $(k+1)^p \equiv k^p + \binom{p-1}{1}k^{p-2} + \cdots + \binom{p-1}{p-2}k + 1\ (\mathrm{mod}\ p)$ $\equiv k + \binom{p}{1}k^{p-1} + \cdots + 1\ (\mathrm{mod}\ p)$이다. (귀류법가정에 의해서) ······ ⑤ 정리 4.25에 의하여 $i < p$이면 $p \mid \binom{p}{i}$ 이다. ······ ⑥ 귀납법에 의해, 모든 자연수 a에 대해 $a^p \equiv a\ (\mathrm{mod}\ p)$이다. ······ ⑦ 그러면 $p \mid a(a^p - 1)$이다. ······ ⑧ $(a,p) = 1$이므로 $p \mid a^p - 1$이고 따라서 $a^p \equiv 1\ (\mathrm{mod}\ p)$이다. ······ ⑨
논리성: 1 / 3 / 5점 이유: **3점**. 5번 문장의 논리가 생각보다 자명하지 않은 것 같습니다. 귀납법 가정이

어떤 식으로 이용되었는지 구체적으로 적어주어야 합니다. 다른 부분에는 논리적인 비약이 없는 것 같습니다.

명료성: 1 / 2 / 3점
이유: **2점**. 1번 문장과 같이 쓰는 것 보다는 '정리 4.16을 이용한다' 와 같은 표현이 더 좋은 것 같습니다. 그리고 5번 문장에서 '귀류법 가정'을 '귀납법 가정'이라고 고쳐주어야 할 것 같습니다.

참신성: 1 / 2점
이유: **2점**. 기존에 잘 알려진 페르마 정리의 증명과는 다르게 귀납법을 적용하였습니다. 새로운 아이디어를 제시했다는 점에서 참신하다고 생각합니다.

증명 5)

Euler's Theorem에 의해 $a^{\phi(p)} \equiv 1 \pmod{p}$가 성립한다. ······ ①
이때, p의 정의에 의해 p는 p보다 작은 어떤 자연수들과 서로소이므로 ······ ②
$\phi(p) = p-1$이 되어 ······ ③
$a^{p-1} \equiv 1 \pmod{p}$가 성립한다. ······ ④

논리성: 1 / 3 / 5점
이유: **5점**. 각 문장 사이에 논리적인 비약이나 오류가 보이지 않습니다.

명료성: 1 / 2 / 3점
이유: **2점**. Euler phi function이 문제에 언급되어 있지 않지만 증명에는 핵심적으로 사용되고 있습니다. $\phi(p)$가 어떠한 함수인지 정의 정도만 적어주는게 좋을 것 같습니다.

참신성: 1 / 2점
이유: 1점. 위 방법은 페르마 정리보다 강력한 정리에서 특수한 경우만 관찰을 한 것입니다. 새로운 아이디어를 제공하지 않았습니다.

ABSTRACT

A study on reliability and validity of online proof peer assessment

Oh Yaerin
Department of Mathematics Education
The Graduate School
Seoul National University

Despite of the importance of proof learning, research has been reported that not only secondary school students but also undergraduate students have difficulties on learning proofs. Especially in the mathematics major course, therefore, we can consider new tool for proof learning. Peer assessment is an activity students assess each others' work. Peer assessment has been applied in university courses mainly on writing task or practical task. It has been found that peer assessment improves teaching-learning and has various effects in the level of department and university as well. On the other hand, there are few study which use peer assessment in undergraduate mathematics courses.

Even though there are various benefit of peer assessment, why peer assessment is not used often in undergraduate mathematics courses? The main reason would be that we are not convinced that peer assessment results are reliable because students are lack of background knowledge and assessment experience. By the way, no research has been explored reliability and validity of proof peer assessment. In this research, thus, we explore

reliability and validity of peer assessment.

In the <Number Theory> course of a university in Seoul, peer assessment has been administered as a weekly assignment for preview. Each week, students were asked to construct proofs and submit them before learning the mathematical contents and four theorems among them were chosen for peer assessment. Students assessed proofs constructed by three different peers of each theorem. While assessing peers' proofs, students scored logical reasoning, clarity and novelty of the proofs following the rubric given by the professor. All the process of peer assessment including submitting proofs, matching assessors and assessees, assessing proofs and returning the assessment results to the assessee has been worked on an online peer assessment system called *Classrpep*.

First, reliability of proof peer assessment has been discovered based on the results of proof peer assessment by the students. A reliability of an assessment can be explored in three different ways—equivalence, consistency and stability. Research data of this study includes results by only one group of students, so stability cannot be examined in this research. To explore reliability of proof peer assessment, therefore, only equivalence and the reliability have been examined.

Second, validity of proof peer assessment has been discovered. To examine validity of peer assessment, we check whether students assessment is similar to expert assessment. To check validity, the researcher designed 'Proof Assessing Task 1' and 'Proof Assessing Task 2' and asked students and experts to perform the tasks.

As a result, the equivalence has been quite high but the consistency has been low. High equivalence means the peer assessment result is similar regardless of assessors. Thus we can dispel worries that the result might be different depending on assessors. Low consistency, meanwhile, indicates the peer assessment scores of each students across the semester is not consistent

but changing. Considering that students are learning proving during the semester, consistency of peer assessment is not something critical to the reliability of the assessment. Thus we cannot conclude low consistency means low reliability of the peer assessment.

According to the analysis of student assessment and expert assessment results, students' assessing performance on logical reasoning was similar to the experts' but students' assessing performance on clarity and novelty was very different to the experts' in the middle of the semester. After the semester ends, students' assessing performance on all of the three criteria—logical reasoning, clarity and novelty—was similar to the experts' assessing performance. This results show that the validity of the proof peer assessment was low at the beginning and the middle of the semester, but the validity became high at the end of the semester. Since students became to assess similar to the experts as time goes by, we could conjecture that students'perspective on proof became more similar to the experts' perspective.

Keywords: proof, peer assessment, reliability, validity

Student Number: **2015-23057**